

Smith, Mona

79815

From: ~~Schreiber, David~~
Sent: Monday, November 04, 2002 4:41 PM
To: Smith, Mona
Cc: Nashed, Nashaat
Subject: 09/724,8746

AV 1652

10001
mailbox
please

Mona,

Please run the following fragments in the rush search for this case.

Seq ID no. 2, residues 16975-17065
residues 18300-18400
residues 19950-20050
residues 20490-20590
residues 21350-21450
residues 22750-22850
residues 24300-24400
residues 24950-25050
residues 25850-25950
residues 27200-27300
residues 28200-28300
residues 29800-29900
residues 30550-30650
residues 31170-31270
residues 32070-32170
residues 33500-33600
residues 34350-34450
residues 36150-36250
residues 36940-37040
residues 37570-37670

Point of Contact:
Mona Smith
Technical Information Specialist
CM1 6A01
Tel: 308 3278

David Schreiber, Ph.D.
Scientific and Technical Information Center
Biotech/Chem Library
CM1-6A03
703-308-4292

Searcher: M. Smith
Phone:
Location:
Date Picked Up: 11/5/02
Date Completed: 11/8/02
Searcher Prep/Review: 15
Clerical:
Online Time 25

NA sequences: 20
AA sequences:
Structures:
Bibliographic
Litigation:
Full text:
Patent Family
Other:

Vendor Cost:
STN:
Dialog:
Orbit/Questel:
Lexis/Nexis:
Sequence Sys:
WWW/internet:
Other:

Point of Contact
Mona Smith
Technical Information Specialist
CMJ 8A01
Tel 808 3058

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 6, 2002, 11:51:54 ; Search time 1197.62 Seconds
(without alignments)
1365.823 Million cell updates/sec

Title: US-09-724-876-2_COPY_37570_37670

Perfect score: 101

Sequence: 1 aggtctgcagctgtctgcgc.....tgggctcagagctgtgcgaac 101

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: EST:
2: em_estdb:
3: em_esthm:
4: em_estln:
5: em_estmu:
6: em_estov:
7: em_estpl:
8: em_estro:
9: em_hlc:
10: gb_est1:
11: gb_est2:
12: gb_hlc:
13: gb_est3:
14: gb_est4:
15: gb_est5:
16: em_estfun:
17: em_estom:
18: gb_gss:
19: em_gss_hum:
20: em_gss_inv:
21: em_gss_pln:
22: em_gss_vrt:
23: em_gss_fun:
24: em_gss_mam:
25: em_gss_mus:
26: em_gss_other:
27: em_gss_pro:
28: em_gss_rtd:
29: em_gss_rtd:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Match Length	ID	Description
1	36.2	35.8	620	BH385803	BH385803 AG-ND-138
2	36.2	35.8	629	BH372094	BH372094 AG-ND-138
3	35.2	34.9	419	AW922099	AW922099 LG1_236_G
4	35.2	34.9	568	BE59451	BE59451 P11_87_F0
5	34.2	33.9	465	BE595103	BE595103 P11_45_F0
6	33.6	33.3	219	AW679204	AW679204 WSI_23_G0

7	33.6	33.3	384	10	BE363648	BE363648 WSI_64_G1
8	33.6	33.3	384	14	BO283127	BO283127 WHE3086_D
9	33.6	33.3	401	10	AW287455	AW287455 WSI_228_B
10	33.6	33.3	402	10	BE363414	BE363414 WSI_62_F0
11	33.6	33.3	404	10	AW564133	AW564133 LG1_282_D
12	33.6	33.3	434	10	BE363488	BE363488 WSI_63_G1
13	33.6	33.3	440	10	AW746099	AW746099 WSI_39_H0
14	33.6	33.3	453	10	AW745893	AW745893 WSI_38_C0
15	33.6	33.3	453	10	BE366824	BE366824 P11_41_A0
16	33.6	33.3	456	13	BC933006	BC933006 WSI_3_A04
17	33.6	33.3	456	13	BC933007	BC933007 WSI_3_A06
18	33.6	33.3	460	10	BE593071	BE593071 WSI_97_B0
19	33.6	33.3	464	10	BE594219	BE594219 WSI_103_G
20	33.6	33.3	467	10	AW747205	AW747205 WSI_66_C0
21	33.6	33.3	468	10	AW678238	AW678238 WSI_14_H0
22	33.6	33.3	468	10	BE597494	BE597494 P11_70_C0
23	33.6	33.3	468	10	BE599472	BE599472 P11_88_A0
24	33.6	33.3	473	10	AW746220	AW746220 WSI_40_D0
25	33.6	33.3	474	10	BE366769	BE366769 P11_41_A0
26	33.6	33.3	475	10	AW677810	AW677810 WSI_11_B0
27	33.6	33.3	475	10	BE364060	BE364060 P11_11_C1
28	33.6	33.3	476	10	AW746140	AW746140 WSI_39_H1
29	33.6	33.3	476	13	BM318387	BM318387 P11_13_A0
30	33.6	33.3	480	13	BM323992	BM323992 P1C1_30_A
31	33.6	33.3	490	13	BM328506	BM328506 P1C1_30_A
32	33.6	33.3	492	10	BE592512	BE592512 WSI_94_A0
33	33.6	33.3	503	10	BE594053	BE594053 WSI_101_F
34	33.6	33.3	504	10	BE593619	BE593619 WSI_98_H0
35	33.6	33.3	511	10	AW745041	AW745041 LG1_386_A
36	33.6	33.3	511	10	AW746104	AW746104 WSI_39_H1
37	33.6	33.3	524	10	AW745368	AW745368 WSI_34_A0
38	33.6	33.3	525	10	AW284481	AW284481 LG1_284_A
39	33.6	33.3	525	10	AW747678	AW747678 WSI_75_B0
40	33.6	33.3	527	10	BE599116	BE599116 P11_85_E0
41	33.6	33.3	527	13	BM326393	BM326393 P1C1_56_E
42	33.6	33.3	528	10	AW745988	AW745988 WSI_38_C0
43	33.6	33.3	532	10	AW746329	AW746329 WSI_40_D0
44	33.6	33.3	533	10	AW680220	AW680220 WSI_50_H0
45	33.6	33.3	538	10	AW746131	AW746131 WSI_39_H0

ALIGNMENTS

RESULT 1
BH385803
LOCUS
DEFINITION AG-ND-13813.TF.1 ND-TAM Anopheles gambiae genomic clone AG-ND-13813
, DNA sequence.
ACCESSION BH385803
VERSION BH385803.1 GI:17331945
KEYWORDS
SOURCE
ORGANISM
Anopheles gambiae
African malaria mosquito.
Anopheles gambiae
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
Amophel.
REFERENCE
1 (bases 1 to 620)
Shetty,J., Malek,J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J.
AUTHORS
Direct Submission of BAC-end sequences from Anopheles gambiae
JOURNAL
Unpublished (2001)
COMMENT
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: b.loftus@tigr.org

This clone is from an A. gambiae BAC library (ND-TAM) provided by F.H. Collins and sequenced by The Institute for Genomic Research (TIGR). The BAC library was generated from A. gambiae PEST strain DNA. All DNA was extracted from newly hatched first instar larvae to minimize the inclusion of DNA from microorganisms that inhabit

the gut. The DNA is derived from mixed sexes of larvae. The BAC library was constructed at Texas A&M University BAC Center University, College Station, Texas 77843-2123, USA using a HindIII partial digest.
Seq primer: M13 For
Class: BAC ends.

FEATURES
source
1..620
/organism="Anopheles gambiae"
/strain="PEST"
/db_xref="taxon:7165"
/clone_1lb="ND-13813"
/note="Vector: pECBAC1; Site_1: HindIII"
BASE COUNT 112 a 211 c 200 g 97 t
ORIGIN

Query Match 35.8%; Score 36.2; DB 17; Length 620;

Best Local Similarity 65.4%; Pred. No. 6.1;

Matches 53; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

QY 8 GCAGGTCGCGCCGCTCCCGAAGCAAGCTGAGTGCCTCCGCTCAGAGCCCTGG 67

Db 276 GAAGCTGCTGCGCCGCTCAGCAAGCGAGCGGCTCTGCGCCCGCATGGGCTGGC 335

QY 68 AATGACTCGCTGATGGGCT 88

Db 336 CATGACGCTCTGCACAGCT 356

RESULT 2

BH372094 629 bp DNA linear GSS 10-DEC-2001

LOCUS AG-ND-13813.TF ND-TAM Anopheles gambiae genomic clone AG-ND-13813,

DEFINITION DNA sequence.

ACCESSION BH372094 GI:17318219

VERSION GSS.

KEYWORDS African malaria mosquito.

SOURCE Anopheles gambiae

ORGANISM Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;

Anopheles.

1 (bases 1 to 629)

REFERENCE Shetty,J., Malek,J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J.

AUTHORS Direct Submission of BAC-end sequences from Anopheles gambiae

JOURNAL Unpublished (2001)

COMMENT Other_GSSs: AG-ND-13813.TR

Contact: Brendan J Loftus

Department of Eukaryotic Genomics

The Institute for Genomic Research

7912 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0208

Fax: 301 838 3543

Email: bjoftus@tigr.org

This clone is from an A. gambiae BAC library (ND-TAM) provided by

F.H. Collins and sequenced by The Institute for Genomic Research

(TIGR). The BAC library was generated from A. gambiae PEST strain

DNA. All DNA was extracted from newly hatched first instar larvae

to minimize the inclusion of DNA from microorganisms that inhabit

the gut. The DNA is derived from mixed sexes of larvae. The BAC

library was constructed at Texas A&M University BAC Center

University, College Station, Texas 77843-2123, USA using a HindIII

partial digest.

Seg primer: M13 For

Class: BAC ends.

Location/Qualifiers

1..629

/organism="Anopheles gambiae"

BASE COUNT 110 a /note="Vector: pECBAC1; Site_1: HindIII"

ORIGIN 209 c 206 g 104 t

Query Match 35.8%; Score 36.2; DB 17; Length 629;

Best Local Similarity 65.4%; Pred. No. 6.1;

Matches 53; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

QY 8 GCAGGTCGCGCCGCTCCCGAAGCAAGCTGAGTGCCTCCGCTCAGAGCCCTGG 67

Db 276 GAAGCTGCTGCGCCGCTCAGCAAGCGAGCGGCTCTGCGCCCGCATGGGCTGGC 335

QY 68 AATGACTCGCTGATGGGCT 88

Db 336 CATGACGCTCTGCACAGCT 356

RESULT 3

AM922099 419 bp mRNA linear EST 19-JUL-2000

LOCUS LGL_236.G05.b1_A002 Light Grown 1 (LGL) Sorghum bicolor cDNA, mRNA

DEFINITION sequence.

ACCESSION AM922099 GI:8087924

VERSION AM922099.1

KEYWORDS EST.

SOURCE sorghum.

ORGANISM Sorghum bicolor

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC

clade; Panicoidae; Andropogoneae; Sorghum.

1 (bases 1 to 419)

REFERENCE Cordomier-Pratt,M.-M., Gingle,A., Marsala,C. and Pratt,L.H.

AUTHORS An EST database from Sorghum: light-grown seedlings

JOURNAL Unpublished (2000)

COMMENT Contact: Cordomier-Pratt MM

Laboratory for Genomics and Bioinformatics

The University of Georgia, Department of Plant Biology

Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA

Tel: 706 542 1860

Fax: 706 583 0210

Email: mmp@trattuga.edu

Sequences have been trimmed to exclude polyA, vector and regions

below Phred quality 16. The threshold for highest quality sequence

is 20.

Seq primer: JEN REV

High quality sequence stop: 208

POLYA-No.

FEATURES

source

1..419

/organism="Sorghum bicolor"

/db_xref="taxon:4558"

/clone_1lb="Light Grown 1 (LGL)"

/note="Organ: 10- to 14-day-old light-grown (greenhouse)

seedlings; Vector: Lambda zap; Site_1: XhoI; Site_2: EcoRI

; The library was made from poly-A RNA in the cloning

vector lambda ZAP II. Clones to be sequenced were

prepared by mass excision."

BASE COUNT 76 a 130 c 149 g 64 t

ORIGIN

Query Match 34.9%; Score 35.2; DB 10; Length 419;

Best Local Similarity 65.0%; Pred. No. 10;

Matches 52; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

QY 22 TCCCGAAGCAAGCTGAGTGCCTCCGCTCAGAGCCCTGGAAATGACTGCTGA 81

Db 317 TCCCGAAGCAAGCTGAGTGCCTCCGCTCAGAGCCCTGGAAATGACTGCTGA 376

QY 82 TGGGCTAGAGCTGCGCAAC 101

Db 377 AGAGGACTACTGCTGGCTAC 396

RESULT 4
 BE599451
 LOCUS
 DEFINITION P11_87_F08.b1_A002 Pathogen induced 1 (P11) Sorghum bicolor cDNA,
 mRNA sequence.
 ACCESSION BE599451
 VERSION BE599451
 KEYWORDS
 SOURCE
 ORGANISM
 Sorghum bicolor
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
 clade; Panicoideae; Andropogoneae; Sorghum.
 1 (bases 1 to 568)
 Cordonnier-Pratt,M.-M., Gingle,A., Dean,R., Sudman,M. and Pratt
 /L.H.
 REFERENCE An EST database from Sorghum: pathogen-induced plants
 AUTHORS Unpublished (2000)
 JOURNAL Contact: Cordonnier-Pratt MM
 COMMENT Laboratory for Genomics and Bioinformatics
 The University of Georgia, Department of Plant Biology
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
 Tel: 706 542 1860
 Fax: 706 583 0210
 Email: mmp@prattuga.edu
 Sequences have been trimmed to exclude PolyA, vector and regions
 below Phred quality 16. The threshold for highest quality sequence
 is 20.
 Seq primer: JEN REV
 High quality sequence stop: 490
 POLYA-No.

FEATURES
 source
 Location/Qualifiers
 1..568
 /organism="Sorghum bicolor"
 /db_xref="taxon:4558"
 /clone_lib="Pathogen induced 1 (P11)"
 /note="Organ: Anthracnose-infected leaves from
 two-week-old sorghum plants 48 hr after inoculation;
 Vector: pBluescript II from lambda Zap II; Site_1: XhoI;
 Site_2: EcoRI; Two-week-old sorghum plants (BTX 623
 cultivar) were infected with pathogen (isolate PRM421 of
 Colletotrichum graminicola, which is a sorghum isolate).
 RNA was prepared from infected leaves harvested from 45
 seedlings 48 hours after inoculation. Note: young
 seedlings (2 weeks old) exhibit juvenile resistant
 reaction, which is an incompatible interaction. As they
 grow older (4 weeks or older), plants resume susceptibility
 to anthracnose disease. The library was made from poly-A
 RNA in the cloning vector lambda Zap II. Clones to be
 sequenced were prepared by mass excision. WARNING: While
 most or all ESTs are expected to derive from the host
 plant, no effort was made to eliminate ESTs deriving from
 the pathogen."

BASE COUNT 103 a 175 c 194 g 96 t
 ORIGIN
 Query Match 34.9%; Score 35.2; DB 10; Length 568;
 Best Local Similarity 65.0%; Pred. No. 11;
 Matches 52; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

QY 22 TCCCGAAGGCAAGCTCGACGTGATGCGCGCTCAGAGCTGGGAATGACTGCTGA 81
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 285 TCCCGAGAGCCCAAGCTCCCGCGCGCGCGCGAGGCGCAGCATGAGGTGACGCGGTGG 344
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 82 TGGGGCTAGAGCTGGCGCAAC 101
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 345 AGAGGACTACTGCTGCTAC 364
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 5
 BE595103
 LOCUS
 DEFINITION P11_45_F08.b1_A002 Pathogen induced 1 (P11) Sorghum bicolor cDNA,
 mRNA sequence.
 ACCESSION BE595103
 VERSION BE595103
 KEYWORDS
 SOURCE
 ORGANISM
 Sorghum bicolor
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
 clade; Panicoideae; Andropogoneae; Sorghum.
 1 (bases 1 to 465)
 Cordonnier-Pratt,M.-M., Gingle,A., Dean,R., Sudman,M. and Pratt
 /L.H.
 REFERENCE An EST database from Sorghum: pathogen-induced plants
 AUTHORS Unpublished (2000)
 JOURNAL Contact: Cordonnier-Pratt MM
 COMMENT Laboratory for Genomics and Bioinformatics
 The University of Georgia, Department of Plant Biology
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
 Tel: 706 542 1860
 Fax: 706 583 0210
 Email: mmp@prattuga.edu
 Sequences have been trimmed to exclude PolyA, vector and regions
 below Phred quality 16. The threshold for highest quality sequence
 is 20.
 Seq primer: JEN REV
 High quality sequence stop: 342
 POLYA-No.

FEATURES
 source
 Location/Qualifiers
 1..465
 /organism="Sorghum bicolor"
 /db_xref="taxon:4558"
 /clone_lib="Pathogen induced 1 (P11)"
 /note="Organ: Anthracnose-infected leaves from
 two-week-old sorghum plants 48 hr after inoculation;
 Vector: pBluescript II from lambda Zap II; Site_1: XhoI;
 Site_2: EcoRI; Two-week-old sorghum plants (BTX 623
 cultivar) were infected with pathogen (isolate PRM421 of
 Colletotrichum graminicola, which is a sorghum isolate).
 RNA was prepared from infected leaves harvested from 45
 seedlings 48 hours after inoculation. Note: young
 seedlings (2 weeks old) exhibit juvenile resistant
 reaction, which is an incompatible interaction. As they
 grow older (4 weeks or older), plants resume susceptibility
 to anthracnose disease. The library was made from poly-A
 RNA in the cloning vector lambda Zap II. Clones to be
 sequenced were prepared by mass excision. WARNING: While
 most or all ESTs are expected to derive from the host
 plant, no effort was made to eliminate ESTs deriving from
 the pathogen."

BASE COUNT 79 a 145 c 169 g 70 t 2 others
 ORIGIN
 Query Match 33.9%; Score 34.2; DB 10; Length 465;
 Best Local Similarity 63.8%; Pred. No. 19;
 Matches 51; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 22 TCCCGAAGGCAAGCTCGACGTGATGCGCGCTCAGAGCTGGGAATGACTGCTGA 81
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 307 TCCCGAGAGCCCAAGCTCCCGCGCGCGCGCGAGGCGCAGCANGAGGTGACGCGGTGG 366
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 82 TGGGGCTAGAGCTGGCGCAAC 101
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 367 AGAGGACTACTGCTGCTAC 386
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 6
 AM679204
 LOCUS
 DEFINITION WS1_23_G01.b1_A002 Water-stressed 1 (WS1) Sorghum bicolor cDNA,
 mRNA sequence.
 ACCESSION AM679204
 VERSION AM679204.1 GI:7552958
 KEYWORDS

SOURCE	sorghum.	
ORGANISM	Sorghum bicolor	
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Sorghum.	
AUTHORS	1 (bases 1 to 219) Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt ,L.H.	
TITLE	An EST database from Sorghum: water-stressed plants	
JOURNAL	Unpublished (2000)	
COMMENT	Contact: Cordonnier-Pratt MM Laboratory for Genomics and Bioinformatics The University of Georgia, Department of Plant Biology Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA Tel: 706 542 1860 Fax: 706 583 0210 Email: mmpratt@uga.edu Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20. Seq primer: JEN REV High quality sequence stop: 218 POLYA-No.	
FEATURES	Location/Qualifiers	
SOURCE	1..219 /organism="Sorghum bicolor" /db_xref="taxon:4558" /clone_id="Water-stressed 1 (WS1)" /note="Organ: Mix of 5-week old plants on days 7 & 8 after water was withheld; Vector: lambda zap; Site.1: XhoI; Site.2: EcoRI; The library was made from poly-A RNA in the cloning vector lambda ZAP II. Clones to be sequenced were prepared by mass excision."	
BASE COUNT	35 a 65 c 88 g 31 t	
ORIGIN		
Query Match	33.3%; Score 33.6; DB 10; Length 219;	
Best Local Similarity	63.8%; Pred. No. 23;	
Matches	51: Conservative 0; Mismatches 29; Indels 0; Gaps 0;	
QY	22 TCCCGAAGCGCAAGCTCGATGATGGCCGCTCACGAGCCTGGGAATGCACTGCTGA 81 93 TCCCGAAGCGCAAGCTCGCGCGCGCGCGCGAGCGACGAGAGGTGACGCGGTGG 152 QY 82 TGGGCTAGAGCTGCCAAC 101 Db 153 AGAGGAGTACTCTGTCTTAC 172	
RESULT 7		
LOCUS	BE363648 384 bp mRNA linear EST 20-JUL-2000	
DEFINITION	WS1_64.G12.g1.A002 Water-stressed 1 (WS1) sorghum bicolor cDNA, mRNA sequence.	
ACCESSION	BE363648	
VERSION	BE363648.1 GI:9305205	
KEYWORDS	EST.	
SOURCE	sorghum.	
ORGANISM	Sorghum bicolor	
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Sorghum.	
AUTHORS	1 (bases 1 to 384) Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt ,L.H.	
TITLE	An EST database from Sorghum: water-stressed plants	
JOURNAL	Unpublished (2000)	
COMMENT	Contact: Cordonnier-Pratt MM Laboratory for Genomics and Bioinformatics The University of Georgia, Department of Plant Biology Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA Tel: 706 542 1860 Fax: 706 583 0210	

Email: mmprratt@uga.edu
Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20.
Seq primer: PolyTmIX
High quality sequence start: 5
High quality sequence stop: 362
POLYA=Yes

FEATURES
Source Location/Qualifiers
1..384
/organism="Sorghum bicolor"
/db_xref="taxon:4558"
/clone_lib="Water-stressed 1 (WS1)"
/note="Organ: Mix of 5-week old plants on days 7 & 8 after water was withheld; Vector: Lambda Zap; Site_1: XhoI; Site_2: EcoRI; The library was made from poly-A RNA in the cloning vector lambda Zap II. Clones to be sequenced were prepared by mass excision."

BASE COUNT
74 a 114 c 129 g 67 t

ORIGIN

Query Match 33.3%; Score 33.6; DB 10; Length 384;
Best Local Similarity 63.8%; Pred. No. 27;
Matches 51; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

OY 22 TCCCGGAAGCAGCTGCAGCTGATGCGCGGCTACAGAGCTGGAGATGCATCTCGTGA 81
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 95 TCCCGAGAGCCAGACGTCCCGCGCGGCGCGGAGGAGCAGAGAGGCTGAGCGCGGTGG 154
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

OY 82 TGGGCTGAGAGCTGCGCAAC 101
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 155 AGAGGGACTACTGTCTCTAC 174
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 8
BQ283127 384 bp mRNA linear EST 13-MAY-2002
LOCUS BQ283127
DEFINITION WHE3086_D11_H22S wheat cold-stressed seedling subtracted cDNA
library Triticum aestivum cDNA clone WHE3086_D11_H22, mRNA
sequence.
ACCESSION BQ283127
VERSION BQ283127.1 GI:20552527
KEYWORDS EST.
SOURCE bread wheat.
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidaeae
; Triticeae; Triticum.
1 (bases 1 to 384)
Anderson,O.D., Chao,S., Choi,D.W., Close,T.J., Crossman,C., Fenton
,R.D., Lazo,G.R., Nguyen,H.T., Pham,J., Rausch,C.J., Wilson,C., Woo
,J. and Zhang,D.
The structure and function of the expressed portion of the wheat
genomes - Cold-stressed seedling subtracted cDNA library
Unpublished (2002)
Contact: Olin Anderson
US Department of Agriculture, Agriculture Research Service, Pacific
West Area, Western Regional Research Center
800 Buchanan Street, Albany, CA 94710, USA
Tel: 5105595773
Fax: 5105595818
Email: oanderson@w.usda.gov
Sequences have been trimmed to remove vector sequence and low
quality sequence with phred score less than 20
Seq primer: SK primer.

FEATURES
Source Location/Qualifiers
1..384
/organism="Triticum aestivum"
/cultivar="Chinese Spring"
/db_xref="taxon:4565"
/clone="WHE3086_D11_H22"
/clone_lib="Wheat cold-stressed seedling subtracted cDNA
library"

```

/organism="Sorghum bicolor"
/db_xref="taxon:4558"
/clone_1ib="Light Grown 1 (LG1)"
/note="Organ: 10- to 14-day-old light-grown (greenhouse)"

```

0y 22 TCCCGGAAGCAGCTCGAGCTGGATGCGCCGCTCACGAGCCTGGCAATGCACTCGCTGA 81
 ||| | | ||||| | | ||||| ||||| | | ||||| | |
 Db 92 TCCCGGAGGCAGCTCGCGCGCGCGCGCGCGCGGAGCGCAGCAGAGGTTGACGCGGTTGG 151

QY 82 TGGGCTAGAGCTGCCAAC 101
 Db 152 AGAGGAGCTACTGTCTTAC 171

RESULT 11
 AM564133 404 bp mRNA linear EST 19-JUL-2000
 LOCUS L61_282.D02.b1_A002 Light Grown 1 (LGI) Sorghum bicolor cDNA, mRNA
 DEFINITION sequence.
 ACCESSION AM564133
 VERSION AM564133.1 GI:7218011
 KEYWORDS EST.
 SOURCE sorghum.
 ORGANISM Sorghum bicolor
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
 clade; Panicoideae; Andropogoneae; Sorghum.
 1 (bases 1 to 404)
 Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C. and Pratt,L.H.
 An EST database from Sorghum: light-grown seedlings
 Unpublished (2000)
 Contact: Cordonnier-Pratt MM
 Laboratory for Genomics and Bioinformatics
 The University of Georgia, Department of Plant Biology
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
 Tel: 706 542 1860
 Fax: 706 583 0210
 Email: mmp@prattuga.edu
 Sequences have been trimmed to exclude PolyA, vector and regions
 below Phred quality 16. The threshold for highest quality sequence
 is 20.

Seq primer: JEN REV
 High quality sequence stop: 403
 POLYA-No.

FEATURES
 source Location/Qualifiers

1..404
 /organism="Sorghum bicolor"
 /db_xref="taxon:4558"
 /clone_lib="Light Grown 1 (LGI)"
 /note="Organ: 10- to 14-day-old light-grown (greenhouse)
 seedlings; Vector: Lambda Zap; Site_1: XhoI; Site_2: EcoRI
 ; The library was made from poly-A RNA in the cloning
 vector Lambda Zap II. Clones to be sequenced were
 prepared by mass excision."
 BASE COUNT 71 a 127 c 144 g 62 t
 ORIGIN

Query Match 33.3%; Score 33.6; DB 10; Length 404;
 Best Local Similarity 63.8%; Pred. No. 27;
 Matches 51; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 22 TCCCGGAGGCAAGCTGCAGTGGCGCGCTCACAGACCTGGGAATGACTCGCTGA 81
 Db 316 TCCCGGAGGCAAGCTGCAGTGGCGCGCGCTCACAGACGAGGCTGACGCGGTGG 375

QY 82 TGGGCTAGAGCTGCCAAC 101
 Db 376 AGAGGAGCTACTGTCTTAC 395

RESULT 12
 BE363488 434 bp mRNA linear EST 20-JUL-2000
 LOCUS WS1_63.G11.g1_A002 Water-stressed 1 (WS1) Sorghum bicolor cDNA,
 DEFINITION mRNA sequence.
 ACCESSION BE363488
 VERSION BE363488.1 GI:9305045
 KEYWORDS EST.
 SOURCE sorghum.
 ORGANISM Sorghum bicolor
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC

REFERENCE
 AUTHORS Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
 ,L.H.
 TITLE An EST database from Sorghum: water-stressed plants
 JOURNAL Unpublished (2000)
 COMMENT Contact: Cordonnier-Pratt MM
 Laboratory for Genomics and Bioinformatics
 The University of Georgia, Department of Plant Biology
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
 Tel: 706 542 1860
 Fax: 706 583 0210
 Email: mmp@prattuga.edu
 Sequences have been trimmed to exclude PolyA, vector and regions
 below Phred quality 16. The threshold for highest quality sequence
 is 20.

Seq primer: PolyTMix
 High quality sequence start: 21
 High quality sequence stop: 434
 POLYA=Yes.

FEATURES
 source Location/Qualifiers

1..434
 /organism="Sorghum bicolor"
 /db_xref="taxon:4558"
 /clone_lib="Water-stressed 1 (WS1)"
 /note="Organ: Mix of 5-week old plants on days 7 & 8 after
 water was withheld; Vector: Lambda Zap; Site_1: XhoI;
 Site_2: EcoRI; The library was made from poly-A RNA in the
 cloning vector Lambda Zap II. Clones to be sequenced were
 prepared by mass excision."
 BASE COUNT 73 a 141 c 157 g 63 t
 ORIGIN

Query Match 33.3%; Score 33.6; DB 10; Length 434;
 Best Local Similarity 63.8%; Pred. No. 28;
 Matches 51; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 22 TCCCGGAGGCAAGCTGCAGTGGCGCGCTCACAGACCTGGGAATGACTCGCTGA 81
 Db 198 TCCCGGAGGCAAGCTGCAGTGGCGCGCGCTCACAGACGAGGCTGACGCGGTGG 257

QY 82 TGGGCTAGAGCTGCCAAC 101
 Db 258 AGAGGAGCTACTGTCTTAC 277

RESULT 13
 AW746099 440 bp mRNA linear EST 19-JUL-2000
 LOCUS WS1_39.H04.b1_A002 Water-stressed 1 (WS1) Sorghum bicolor cDNA,
 DEFINITION mRNA sequence.
 ACCESSION AW746099
 VERSION AW746099.1 GI:7659837
 KEYWORDS EST.
 SOURCE sorghum.
 ORGANISM Sorghum bicolor
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
 clade; Panicoideae; Andropogoneae; Sorghum.
 1 (bases 1 to 440)
 Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
 ,L.H.
 An EST database from Sorghum: water-stressed plants
 Unpublished (2000)
 Contact: Cordonnier-Pratt MM
 Laboratory for Genomics and Bioinformatics
 The University of Georgia, Department of Plant Biology
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
 Tel: 706 542 1860
 Fax: 706 583 0210
 Email: mmp@prattuga.edu
 Sequences have been trimmed to exclude PolyA, vector and regions
 below Phred quality 16. The threshold for highest quality sequence

is 20.
Seq primer: JEN REV
High quality sequence stop: 432
POLYA-No.

FEATURES
Source Location/Qualifiers
1..440

/organism="Sorghum bicolor"
/db_xref="taxon:4558"
/clone_lib="Water-stressed 1 (WS1)"
/note="Organ: Mix of 5-week old plants on days 7 & 8 after water was withheld; Vector: Lambda Zap; Site:1: XhoI; Site:2: EcoRI; The library was made from poly-A RNA in the cloning vector Lambda Zap II. Clones to be sequenced were prepared by mass excision."
BASE COUNT 76 a 140 c 156 g 68 t
ORIGIN

Query Match 33.3%; Score 33.6; DB 10; Length 440;
Best Local Similarity 63.8%; Pred. No. 28;
Matches 51; Conservative 0; Mismatches 29; Indels 0; Gaps 0;
QY 22 TCCCGAGGCAAGCTCGACGTGATGCCGCCGCTACAGCCTGGGATGACCTGCTGA 81
Db 311 TCCCGAGGCAAGCTCGACGTGATGCCGCCGCTACAGCCTGGGATGACCTGCTGA 370
QY 82 TGGGCTAGAGCTGGCGCAAC 101
Db 371 AGAGGACTACCTGTCTAC 390

RESULT 14
LOCUS AM745893 453 bp mRNA linear EST 19-JUL-2000
DEFINITION WS1_38_C06_b1_A002 Water-stressed 1 (WS1) Sorghum bicolor cDNA,
ACCESSION AM745893
VERSION AM745893.1 GI:7659631
KEYWORDS EST.
SOURCE sorghum.
ORGANISM Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE 1 (bases 1 to 453)
AUTHORS Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
L.H.

TITLE An EST database from Sorghum: water-stressed plants
JOURNAL Unpublished (2000)
COMMENT Contact: Cordonnier-Pratt MM
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology,
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: mmp@atuga.edu

Sequences have been trimmed to exclude PolyA, vector and regions
below Phred quality 16. The threshold for highest quality sequence
is 20.
Seq primer: JEN REV
High quality sequence stop: 450
POLYA-No.

FEATURES
Source Location/Qualifiers
1..453

/organism="Sorghum bicolor"
/db_xref="taxon:4558"
/clone_lib="Water-stressed 1 (WS1)"
/note="Organ: Mix of 5-week old plants on days 7 & 8 after water was withheld; Vector: Lambda Zap; Site:1: XhoI; Site:2: EcoRI; The library was made from poly-A RNA in the cloning vector Lambda Zap II. Clones to be sequenced were prepared by mass excision."
BASE COUNT 76 a 142 c 164 g 71 t
ORIGIN

Query Match 33.3%; Score 33.6; DB 10; Length 453;
Best Local Similarity 63.8%; Pred. No. 28;
Matches 51; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 22 TCCCGAGGCAAGCTCGACGTGATGCCGCCGCTACAGCCTGGGATGACCTGCTGA 81
Db 304 TCCCGAGGCAAGCTCGACGTGATGCCGCCGCTACAGCCTGGGATGACCTGCTGA 363
QY 82 TGGGCTAGAGCTGGCGCAAC 101
Db 364 AGAGGACTACCTGTCTAC 383

RESULT 15
LOCUS BE366824 453 bp mRNA linear EST 20-JUL-2000
DEFINITION P11_41_A02_g1_A002 Pathogen induced 1 (P11) Sorghum bicolor cDNA,
ACCESSION BE366824
VERSION BE366824.1 GI:9308381
KEYWORDS EST.
SOURCE sorghum.
ORGANISM Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE 1 (bases 1 to 453)
AUTHORS Cordonnier-Pratt,M.-M., Gingle,A., Dean,R., Sudman,M. and Pratt
L.H.

TITLE An EST database from Sorghum: pathogen-induced plants
JOURNAL Unpublished (2000)
COMMENT Contact: Cordonnier-Pratt MM
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology,
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: mmp@atuga.edu

Sequences have been trimmed to exclude PolyA, vector and regions
below Phred quality 16. The threshold for highest quality sequence
is 20.
Seq primer: POLYTMix
High quality sequence start: 3
High quality sequence stop: 448
POLYA-No.

FEATURES
Source Location/Qualifiers
1..453

/organism="Sorghum bicolor"
/db_xref="taxon:4558"
/clone_lib="Pathogen induced 1 (P11)"
/note="Organ: Anthracnose-infected leaves from two-week-old sorghum plants 48 hr after inoculation; Vector: pBluescript II from Lambda Zap II; Site:1: XhoI; Site:2: EcoRI; Two-week-old sorghum plants (BTX 623 cultivar) were infected with pathogen (isolate FRM421 of Colletotrichum graminiicola, which is a sorghum isolate). RNA was prepared from infected leaves harvested from 45 seedlings (2 weeks old) exhibit juvenile resistant reaction, which is an incompatible interaction. As they grow older (4 weeks or older), plants resume susceptibility to anthracnose disease. The library was made from poly-A RNA in the cloning vector Lambda Zap II. Clones to be sequenced were prepared by mass excision. WARNING: While most or all ESTs are expected to derive from the host plant, no effort was made to eliminate ESTs deriving from the pathogen."
BASE COUNT 104 a 114 c 140 g 95 t
ORIGIN

Query Match 33.3%; Score 33.6; DB 10; Length 453;
Best Local Similarity 63.8%; Pred. No. 28;

	Matches	51;	Conservative	0;	Mismatches	29;	Indels	0;	Gaps	0;
QY	22	TCCCGAAGGCAAGCTGACGTGATGCGCGCTCACGAGCCTGGGAATGACTCGCTGA	81							
Db	15	TCCCGAAGGCAAGCTGCGCGCGCGCGCGGAGGAGGAGCAGAGAGGCTGGACCGGTGG	74							
QY	82	TGGGCTAGAGCTGCGCAC	101							
Db	75	AGAGGACTACTGTCTTAC	94							

Search completed: November 6, 2002, 15:52:51
 Job time : 1200.62 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 6, 2002, 13:32:49 ; Search time 27.125 Seconds
(without alignments)
1240.503 Million cell updates/sec

Title: US-09-724-876-2_COPY_37570_37670

Perfect score: 101
Sequence: 1 agctcgcaggtgctgcgc.....tggggctagagctgcgcac 101

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 310279 seqs, 16657418 residues

Total number of hits satisfying chosen parameters: 620558

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Published Applications -NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
- 10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/2/pubpna/US10_PUB.seq:*
- 12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	30.4	30.1	3209	US-09-925-301-474	Sequence 474, App
2	29.8	29.5	11220	US-09-861-289-32	Sequence 32, Appl
3	29.8	29.5	36778	US-09-861-289-5	Sequence 5, Appl
4	29	28.7	1007	US-09-764-898-58	Sequence 58, Appl
5	28.9	28.7	15872	US-09-861-289-1	Sequence 1, Appl
6	28.8	28.5	290	US-09-294-0938-5935	Sequence 5935, Ap
7	28.8	28.5	2638	US-09-954-456-1955	Sequence 1955, Ap
8	28	27.7	2125	US-09-880-107-2412	Sequence 2412, Ap
9	28	27.7	4149	US-10-044-090-104	Sequence 104, App
10	28	27.7	5086	US-09-880-107-3947	Sequence 3947, App
11	28	27.7	5145	US-09-925-299-206	Sequence 206, App
12	28	27.7	5416	US-09-954-456-786	Sequence 786, App
13	28	27.7	5416	US-09-880-107-2094	Sequence 2094, Ap
14	28	27.3	5432	US-10-044-090-32	Sequence 22, Appl
15	27.6	27.3	245	US-09-884-441-305	Sequence 305, App
16	27.6	27.3	999	US-09-815-242-7966	Sequence 7966, Ap
17	27.6	27.3	1482	US-09-815-242-7954	Sequence 7954, Ap
18	27.2	26.9	1395	US-09-815-242-8008	Sequence 8008, Ap
19	27.2	26.9	2172	US-09-815-242-4038	Sequence 4038, Ap

c	20	27	26.7	419	10	US-09-960-352-14518	Sequence 14518, A
c	21	27	26.7	456	10	US-09-864-761-41	Sequence 41, Appl
c	22	27	26.7	642	10	US-09-864-761-16883	Sequence 16883, A
c	23	27	26.7	2427	10	US-09-254-783A-2	Sequence 2, Appl
c	24	27	26.7	2427	12	US-10-152-058-2	Sequence 26, Appl
c	25	27	26.7	35100	10	US-09-782-378A-26	Sequence 7680, Ap
c	26	26.8	26.5	1503	10	US-09-815-242-7680	Sequence 61, Appl
c	27	26.8	26.5	5681	9	US-09-974-298-61	Sequence 58, Appl
c	28	26.8	26.5	5681	10	US-09-919-1172-58	Sequence 721, App
c	29	26.8	26.5	5594	10	US-09-954-456-1825	Sequence 1825, Ap
c	30	26.8	26.5	5594	10	US-09-954-456-1825	Sequence 782, App
c	31	26.6	26.3	6728	10	US-09-954-456-782	Sequence 3946, App
c	32	26.6	26.3	6728	10	US-09-880-107-3946	Sequence 15673, A
c	33	26.4	26.1	266	10	US-09-876-574-15673	Sequence 3, Appl
c	34	26.4	26.1	1041	10	US-09-915-790-3	Sequence 1, Appl
c	35	26.4	26.1	1275	10	US-09-734-032-1	Sequence 1, Appl
c	36	26.4	26.1	1275	12	US-10-016-985-1	Sequence 1, Appl
c	37	26.4	26.1	1360	10	US-10-016-985-1	Sequence 2269, Ap
c	38	26.4	26.1	1473	12	US-10-016-985-3	Sequence 9, Appl
c	39	26.4	26.1	1902	10	US-09-866-582-9	Sequence 1, Appl
c	40	26.4	26.1	2598	10	US-09-816-094-1	Sequence 1, Appl
c	41	26.4	26.1	2893	10	US-09-916-790-1	Sequence 36, Appl
c	42	26.4	26.1	4041	10	US-09-861-289-36	Sequence 38, Appl
c	43	26.2	25.9	788	10	US-09-861-893-38	Sequence 7713, Ap
c	44	26.2	25.9	1929	10	US-09-815-242-7713	Sequence 3, Appl
c	45	26.2	25.9	7301	10	US-09-816-094-3	

ALIGNMENTS

RESULT 1
US-09-925-301-474
; Sequence 474, Application US/09925301
; Patent No. US20020052308A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA106
; CURRENT APPLICATION NUMBER: US/09/925, 301
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05882
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124, 270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1694
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 474
; LENGTH: 3209
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (427)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-301-474

Query Match 30.1%; Score 30.4; DB 10; Length 3209;
Best Local Similarity 67.2%; Pred. No. 0.87;
Matches 43; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 37 TCACAGTGGATGGCGGCTCAGCAGCGCTGGAGATGACACGCTGAGGCGCTAGACTGC 96
DB 1263 TCACCTGGACAGCTCAGCTGGCGGCTGGCTGACCTGCATGACGCTGAGGCTGC 1322
QY 97 GCAA 100
DB 1323 GCCA 1326

RESULT 2
US-09-861-289-32
; Sequence 32, Application US/09861289

Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600,438US1
; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 32
; LENGTH: 11220
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-861-289-32

Query Match 29.5%; Score 29.8; DB 10; Length 11220;
Best Local Similarity 66.2%; Pred. No. 1.5;
Matches 43; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

OY 37 TCGACGTGATGCGCCGCTCAGAGCCTGGGATGAGTCTCGCTGATGGGGCTAGAGCTGC 96
DB 4400 TCGACAGCGGGCGGCTTCCTGCTGACTCGGATTCGACTCGCTGAGCGGCTGAGCTCC 4459

OY 97 GCAC 101
DB 4460 GCAC 4464

RESULT 3
US-09-861-289-5
; Sequence 5, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600,438US1
; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 36778
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-861-289-5

Query Match 29.5%; Score 29.8; DB 10; Length 36778;
Best Local Similarity 66.2%; Pred. No. 1.6;
Matches 43; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

OY 37 TCGACGTGATGCGCCGCTCAGAGCCTGGGATGAGTCTCGCTGATGGGGCTAGAGCTGC 96
DB 20087 TCGACAGCGGGCGGCTTCCTGCTGACTCGGATTCGACTCGCTGAGCGGCTGAGCTCC 20146

OY 97 GCAC 101
DB 20147 GCAC 20151

RESULT 4
US-09-764-898-58
; Sequence 58, Application US/09764898
; Patent No. US20020090673A1

; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: P201
; CURRENT APPLICATION NUMBER: US/09/764,898
; CURRENT FILING DATE: 2001-01-17
; PRIOR APPLICATION data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 311
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 58
; LENGTH: 1007
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-898-58

Query Match 28.7%; Score 29; DB 10; Length 1007;
Best Local Similarity 57.0%; Pred. No. 2;
Matches 53; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

OY 8 GCAGTGTGCGGCTCCCGGAGGCAAGCTGAGTGGCCGCTCAGAGCTGGG 67
DB 590 GCAGTGTGCGGCTCCCGGAGGCAAGCTGAGTGGCCGCTCAGAGCTGGGCTGGG 649
OY 68 AATGACTGCTGATGGGGCTAGAGCTGCGCA 100
DB 650 CGAGAGCGCTGGCGGCGGCGGCGGCGCA 682

RESULT 5
US-09-861-289-1
; Sequence 1, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600,438US1
; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 15872
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-861-289-1

Query Match 28.7%; Score 29; DB 10; Length 15872;
Best Local Similarity 58.8%; Pred. No. 2.6;
Matches 50; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

OY 17 GCGCTCCCGGAGCAAGCTGACGTGATGCGCCGCTCAGAGCTGGGATGAGTGC 76
DB 8233 GCAGCGGCTGCGGCTGCGGCGGCGGAGCGGCTTCCTCAAGAGCTGCGCTGACTC 8292
OY 77 GCTGATGGGCTAGAGCTGCGCAAC 101
DB 8293 CCGGCGGCGGCTGAGACTCGGCAAC 8317

RESULT 6
US-09-294-093B-5935
; Sequence 5935, Application US/09294093B
; Patent No. US2001005135A1
; GENERAL INFORMATION:
; APPLICANT: Ialjudi, Raghunath, V.
; APPLICANT: Ito, Laura, Y.
; APPLICANT: Sherman, Bradley, K.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN TASSEL


```
; FILE REFERENCE: PL-0009 US
; CURRENT APPLICATION NUMBER: US/09/294,093B
; CURRENT FILING DATE: 1999-04-16
; PRIOR APPLICATION NUMBER: 60/082,567
; PRIOR FILING DATE: April 21, 1998
; NUMBER OF SEQ ID NOS: 6207
; SOFTWARE: PERL Program
; SEQ ID NO 5935
; LENGTH: 290
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: Incyte ID No. US20010051335A1 700382685H1
; NAME/KEY: unsure
; LOCATION: 29
; OTHER INFORMATION: a, t, c, g, or other
US-09-294-093B-5935
```

```
Query Match          28.5%; Score 28.8; DB 10; Length 290;
Best Local Similarity 58.0%; Pred. No. 2.1;
Matches 51; Conservative 0; Mismatches 37; Indels 0; Gaps 0;
```

```
Oy 9 CAGGTGCTGCGGCTCCCGCAAGCAAGCTCGATGCGCGCGCTGACGAGCTGGGA 68
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 40 CTGGAGCTGCTGCTCCCTGCTGCGGAGACCGCGGAGCTTCGCTTACCAAGTTCCTTA 99
Oy 69 ATGACTCGCTGATGGGCTAGAGCTGC 96
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 100 AAGAACTGCTGATGCGGAGAGAGCTCC 127
```

RESULT 7

```
US-09-954-456-1955
; Sequence 1955, Application US/09954456
; Patent No. US20020115057A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Young, Paul
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Canc
; TITLE OF INVENTION: Sets
```

```
; FILE REFERENCE: 689290-76
```

```
; CURRENT APPLICATION NUMBER: US/09/954, 456
```

```
; CURRENT FILING DATE: 2001-09-18
```

```
; PRIOR APPLICATION NUMBER: US/60/233, 617
```

```
; PRIOR FILING DATE: 2000-09-18
```

```
; PRIOR APPLICATION NUMBER: US/60/234, 052
```

```
; PRIOR FILING DATE: 2000-09-20
```

```
; PRIOR APPLICATION NUMBER: US/60/234, 923
```

```
; PRIOR FILING DATE: 2000-09-25
```

```
; PRIOR APPLICATION NUMBER: US/60/235, 134
```

```
; PRIOR FILING DATE: 2000-09-25
```

```
; PRIOR APPLICATION NUMBER: US/60/235, 637
```

```
; PRIOR FILING DATE: 2000-09-26
```

```
; PRIOR APPLICATION NUMBER: US/60/235, 638
```

```
; PRIOR FILING DATE: 2000-09-26
```

```
; PRIOR APPLICATION NUMBER: US/60/235, 711
```

```
; PRIOR FILING DATE: 2000-09-27
```

```
; PRIOR APPLICATION NUMBER: US/60/235, 720
```

```
; PRIOR FILING DATE: 2000-09-27
```

```
; PRIOR APPLICATION NUMBER: US/60/235, 840
```

```
; PRIOR FILING DATE: 2000-09-27
```

```
; PRIOR APPLICATION NUMBER: US/60/235, 863
```

```
; PRIOR FILING DATE: 2000-09-27
```

```
; NUMBER OF SEQ ID NOS: 2276
```

```
; SOFTWARE: PERL Program
```

```
; SEQ ID NO 1955
```

```
; LENGTH: 2638
```

```
; TYPE: DNA
```

```
; ORGANISM: Homo sapiens
```

```
US-09-954-456-1955
```

```
Query Match          28.5%; Score 28.8; DB 10; Length 2638;
Best Local Similarity 58.0%; Pred. No. 2.6;
```

```
Matches 51; Conservative 0; Mismatches 37; Indels 0; Gaps 0;
Oy 9 CAGGTGCTGCGGCTCCCGCAAGCAAGCTCGATGCGCGCGCTGACGAGCTGGGA 68
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2114 CTGGAGCTGCTGCTCCCTGCTGCGGAGACCGCGGAGCTTCGCTTACCAAGTTCCTTA 2173
Oy 69 ATGACTCGCTGATGGGCTAGAGCTGC 96
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2174 AAGAACTGCTGATGCGGAGAGAGCTCC 2201
```

RESULT 8

```
US-09-880-107-2412
; Sequence 2412, Application US/09880107
; Patent No. US20020142981A1
```

```
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
```

```
; APPLICANT: Gene Logic, Inc.
```

```
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
```

```
; FILE REFERENCE: 44921-5028-WO
```

```
; CURRENT APPLICATION NUMBER: US/09/880,107
```

```
; CURRENT FILING DATE: 2001-06-14
```

```
; PRIOR APPLICATION NUMBER: US 60/211,379
```

```
; PRIOR FILING DATE: 2000-06-14
```

```
; PRIOR APPLICATION NUMBER: US 60/237,054
```

```
; PRIOR FILING DATE: 2000-10-02
```

```
; NUMBER OF SEQ ID NOS: 3950
```

```
; SOFTWARE: PatentIn Ver. 2.1
```

```
; SEQ ID NO 2412
```

```
; LENGTH: 2125
```

```
; TYPE: DNA
```

```
; ORGANISM: Homo sapiens
```

```
; FEATURE:
```

```
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 M86826
```

```
US-09-880-107-2412
```

```
Query Match          27.7%; Score 28; DB 10; Length 2125;
Best Local Similarity 55.0%; Pred. No. 4.4;
Matches 55; Conservative 0; Mismatches 45; Indels 0; Gaps 0;
```

```
Oy 2 GGTCTCGAGAGTGTGCGGCTCCCGAAGCAAGCTCGATGCGCGCGCTGACGAG 61
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 926 GGGCTGCTGCTGCTGCGGCTTCCACACACCCATCCGACCTGCGGCGGACCTT 985
Oy 62 CCTGGAAATGACTGCTGATGGGCTAGAGCTGCGCAAC 101
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 986 CAAGAGCTGCACTTCTTGGAGAGAGCTGAGCTGCGGCCAC 1025
```

RESULT 9

```
US-10-044-090-104
; Sequence 104, Application US/10044090
; Patent No. US20020137081A1
```

```
; GENERAL INFORMATION:
```

```
; APPLICANT: Olga Bandman
```

```
; TITLE OF INVENTION: GENES DIFFERENTIALLY EXPRESSED IN VASCULAR TISSUE ACTIVATION
```

```
; FILE REFERENCE: PA-0028 US
```

```
; CURRENT APPLICATION NUMBER: US/10/044,090
```

```
; CURRENT FILING DATE: 2002-01-09
```

```
; NUMBER OF SEQ ID NOS: 850
```

```
; SOFTWARE: PERL Program
```

```
; SEQ ID NO 104
```

```
; LENGTH: 4149
```

```
; TYPE: DNA
```

```
; ORGANISM: Homo sapiens
```

```
; FEATURE:
```

```
; NAME/KEY: misc.feature
```

```
; OTHER INFORMATION: Incyte ID No. US20020137081A1 1383093.19
```

```
; NAME/KEY: unsure
```

```
; LOCATION: 1492-1495, 1663, 1669, 4012, 4055
```

```
; OTHER INFORMATION: a, t, c, g, or other
```

US-10-044-090-104

Query Match 27.7%; Score 28; DB 12; Length 4149;
Best Local Similarity 58.3%; Pred. No. 4.7;
Matches 49; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

QY 13 TCGTGGCCTCCCGAAGCACTGCATGCGCCGCTCAGCAGCCTGGGAATGG 72
DB 1312 TCGTGGTCCCTCGGGAAGCCCTGCTGTAACGTGTGAGTGTGCTGTGCCCCAATGG 1371
QY 73 ACTCGCTGATGGGGCTAGAGCTGC 96
DB 1372 ATTGTGCTGCTGTGCTGTGCTGC 1395

RESULT 10

US-09-880-107-3947
; Sequence 3947, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles In Liver Cancer
; FILE REFERENCE: 44921-5028-MO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3947
; LENGTH: 5086
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 274616
; NAME/KEY: unsure
; LOCATION: (1)..(5086)
; OTHER INFORMATION: n = a or c or g or t
US-09-880-107-3947

Query Match 27.7%; Score 28; DB 10; Length 5086;
Best Local Similarity 58.3%; Pred. No. 4.7;
Matches 49; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

QY 13 TCGTGGCCTCCCGAAGCACTGCATGCGCCGCTCAGCAGCCTGGGAATGG 72
DB 2251 TCGTGGTCCCTCGGGAAGCCCTGCTGTAACGTGTGAGTGTGCTGTGCCCCAATGG 2310
QY 73 ACTCGCTGATGGGGCTAGAGCTGC 96
DB 2311 ATTGTGCTGCTGTGCTGTGCTGC 2334

RESULT 11

US-09-925-299-206
; Sequence 206, Application US/09925299
; Patent No. US20020055627A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 206
; LENGTH: 5145
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (4)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc.feature
; LOCATION: (17)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc.feature
; LOCATION: (5126)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc.feature
; LOCATION: (5143)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-299-206

Query Match 27.7%; Score 28; DB 10; Length 5145;
Best Local Similarity 58.3%; Pred. No. 4.8;
Matches 49; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

QY 13 TCGTGGCCTCCCGAAGCACTGCATGCGCCGCTCAGCAGCCTGGGAATGG 72
DB 2253 TCGTGGTCCCTCGGGAAGCCCTGCTGTAACGTGTGAGTGTGCTGTGCCCCAATGG 2312
QY 73 ACTCGCTGATGGGGCTAGAGCTGC 96
DB 2313 ATTGTGCTGCTGTGCTGTGCTGC 2336

RESULT 12

US-09-954-456-786
; Sequence 786, Application US/09954456
; Patent No. US20020115057A1
; GENERAL INFORMATION:
; APPLICANT: Young, Paul
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using C
; FILE REFERENCE: 689290-76
; CURRENT APPLICATION NUMBER: US/09/954,456
; CURRENT FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: US/60/233,617
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234,052
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,923
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,134
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,637
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,638
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,711
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,720
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,840
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,863
; PRIOR FILING DATE: 2000-09-27
; NUMBER OF SEQ ID NOS: 2276
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 786
; LENGTH: 5416
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-954-456-786

Query Match 27.7%; Score 28; DB 10; Length 5416;

Best Local Similarity 58.3%; Pred. No. 4.8;
Matches 49; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

QY 13 TGTCTGGCTCCCGAAGCAAGCTGACGTGATGCCGCCGCTCAGACCTGGGAATGG 72
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 2581 TGTCTGGCTCCCGAAGCAAGCTGACGTGATGCCGCCGCTCAGACCTGGGAATGG 2640
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 73 ACTCGCTGATGGGCTAGAGCTGC 96
| | | | | | | | | | | | | | | | | | | | | |

Db 2641 ATTTCGTGCTCCGCTGCTGCTGC 2664
| | | | | | | | | | | | | | | | | | | | | |

RESULT 13

US-09-880-107-2094
; Sequence 2094, Application US/09880107
; Patent No. US20020142981A1

GENERAL INFORMATION:

; APPLICANT: Horne, Darci T.

; APPLICANT: Vockley, Joseph G.

; APPLICANT: Scherf, Uwe

; APPLICANT: Gene Logic, Inc.

; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer

; FILE REFERENCE: 44921-5028-WO

; CURRENT APPLICATION NUMBER: US/09/880.107

; CURRENT FILING DATE: 2001-06-14

; PRIOR APPLICATION NUMBER: US 60/211,379

; PRIOR FILING DATE: 2000-06-14

; PRIOR APPLICATION NUMBER: US 60/237,054

; PRIOR FILING DATE: 2000-10-02

; NUMBER OF SEQ ID NOS: 3950

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 2094

; LENGTH: 5416

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; OTHER INFORMATION: Genbank Accession No. US20020142981A1 J03464

US-09-880-107-2094

Query Match 27.7%; Score 28; DB 10; Length 5416;
Best Local Similarity 58.3%; Pred. No. 4.8;

Matches 49; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

QY 13 TGTCTGGCTCCCGAAGCAAGCTGACGTGATGCCGCCGCTCAGACCTGGGAATGG 72
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 2581 TGTCTGGCTCCCGAAGCAAGCTGACGTGATGCCGCCGCTCAGACCTGGGAATGG 2640
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 73 ACTCGCTGATGGGCTAGAGCTGC 96
| | | | | | | | | | | | | | | | | | | | | |

Db 2641 ATTTCGTGCTCCGCTGCTGCTGC 2664
| | | | | | | | | | | | | | | | | | | | | |

RESULT 14

US-10-044-090-22
; Sequence 22, Application US/10044090
; Patent No. US20020137081A1

GENERAL INFORMATION:

; APPLICANT: Olga Bandman

; TITLE OF INVENTION: GENES DIFFERENTIALLY EXPRESSED IN VASCULAR TISSUE ACTIVATION

; FILE REFERENCE: PA-0028 US

; CURRENT APPLICATION NUMBER: US/10/044,090

; CURRENT FILING DATE: 2002-01-09

; NUMBER OF SEQ ID NOS: 850

; SOFTWARE: PERL Program

; SEQ ID NO 22

; LENGTH: 5432

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc_feature

; OTHER INFORMATION: Incyte ID NO. US20020137081A1 1383093.13

US-10-044-090-22

Query Match 27.7%; Score 28; DB 12; Length 5432;
Best Local Similarity 58.3%; Pred. No. 4.8;
Matches 49; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

QY 13 TGTCTGGCTCCCGAAGCAAGCTGACGTGATGCCGCCGCTCAGACCTGGGAATGG 72
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 2584 TGTCTGGCTCCCGAAGCAAGCTGACGTGATGCCGCCGCTCAGACCTGGGAATGG 2643
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 73 ACTCGCTGATGGGCTAGAGCTGC 96
| | | | | | | | | | | | | | | | | | | | | |

Db 2644 ATTTCGTGCTCCGCTGCTGCTGC 2667
| | | | | | | | | | | | | | | | | | | | | |

RESULT 15

US-09-884-441-305/C
; Sequence 305, Application US/09884441
; Patent No. US20020119158A1

GENERAL INFORMATION:

; APPLICANT: Algate, Paul A.

; APPLICANT: Carter, Darrick

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND

; FILE REFERENCE: 210121.462C7

; CURRENT APPLICATION NUMBER: US/09/884,441

; CURRENT FILING DATE: 2001-06-18

; NUMBER OF SEQ ID NOS: 489

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 305

; LENGTH: 245

; TYPE: DNA

; ORGANISM: Homo sapien

; FEATURE:

; NAME/KEY: misc_feature

; LOCATION: (1)...(245)

; OTHER INFORMATION: n = A,T,C or G

US-09-884-441-305

Query Match 27.3%; Score 27.6; DB 10; Length 245;

Best Local Similarity 57.1%; Pred. No. 4.7;

Matches 48; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 13 TGTCTGGCTCCCGAAGCAAGCTGACGTGATGCCGCCGCTCAGACCTGGGAATGG 72
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 176 TGTCTGGCTCCCGAAGCAAGCTGACGTGATGCCGCCGCTCAGACCTGGGAATGG 117
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 73 ACTCGCTGATGGGCTAGAGCTGC 96
| | | | | | | | | | | | | | | | | | | | | |

Db 116 ATTTCGTGCTCCGCTGCTGCTGC 93
| | | | | | | | | | | | | | | | | | | | | |

Search completed: November 6, 2002, 20:33:28
Job time : 44.125 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 6, 2002, 11:51:54 ; Search time 33.5 Seconds
(without alignments)
924.608 Million cell updates/sec

Title: US-09-724-876-2_COPY_37570_37670

Perfect score: 101

Sequence: 1 agctctgcagctgctgcgc.....tggggctagagctgcgcac 101

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: Issued_patents_NA.*
2: /cgn2_6/ptodata/2/ina/5A.COMB.seq.*
3: /cgn2_6/ptodata/2/ina/5B.COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6A.COMB.seq.*
5: /cgn2_6/ptodata/2/ina/6B.COMB.seq.*
6: /cgn2_6/ptodata/2/ina/Backfile1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	101	100.0	71989	US-09-443-501A-2	Sequence 2, Appl1
2	99.4	98.4	68750	US-09-335-409-1	Sequence 1, Appl1
3	99.4	98.4	68750	US-09-568-102-1	Sequence 1, Appl1
4	99.4	98.4	68750	US-09-567-969-1	Sequence 1, Appl1
5	99.4	98.4	68750	US-09-568-480-1	Sequence 1, Appl1
6	99.4	98.4	68750	US-09-568-486-1	Sequence 1, Appl1
7	99.4	98.4	68750	US-09-568-472-1	Sequence 1, Appl1
8	99.4	98.4	68750	US-09-567-899-1	Sequence 1, Appl1
9	35.2	34.9	33529	US-09-144-085-3	Sequence 3, Appl1
10	32.8	32.5	43280	US-08-804-227C-1	Sequence 1, Appl1
11	32.6	32.3	20235	US-07-642-734C-3	Sequence 3, Appl1
12	32.6	32.3	20235	US-08-439-009A-3	Sequence 3, Appl1
13	31.2	30.9	80161	US-09-036-987A-1	Sequence 1, Appl1
14	31.2	30.9	80161	US-09-370-700-1	Sequence 1, Appl1
15	30.4	30.1	8460	US-08-469-005A-9	Sequence 9, Appl1
16	30.4	30.1	8519	US-09-261-907-1	Sequence 1, Appl1
17	30.4	30.1	44377	US-08-804-227C-7	Sequence 7, Appl1
18	30.4	30.1	44377	US-08-804-198-1	Sequence 1, Appl1
19	29.8	29.5	11220	US-09-105-537-5	Sequence 32, Appl1
20	29.8	29.5	36778	US-09-105-537-5	Sequence 32, Appl1
21	29.8	29.5	38506	US-09-320-878-19	Sequence 19, Appl1
22	29.8	29.5	4403765	US-09-103-840A-2	Sequence 2, Appl1
23	29.8	29.5	441529	US-09-103-840A-1	Sequence 1, Appl1
24	29.4	29.1	13987	US-08-804-227C-13	Sequence 13, Appl1
25	29.2	28.9	1734	US-08-804-227C-13	Sequence 13, Appl1
26	29.2	28.9	1734	US-07-642-734C-1	Sequence 1, Appl1
27	29.2	28.7	11219	US-08-439-009A-1	Sequence 1, Appl1

28	29	28.7	15872	US-09-105-537-1	Sequence 1, Appl1
c	29	28.4	328	US-09-060-756-702	Sequence 702, App
c	30	28.4	491	US-09-060-756-107	Sequence 107, App
c	31	28.4	8501	US-08-793-900-1	Sequence 1, Appl1
c	32	27.6	245	US-09-404-879A-305	Sequence 305, App
c	33	27.4	2588	US-08-796-414B-6	Sequence 6, Appl1
c	34	27.2	868	US-09-376-728-3	Sequence 3, Appl1
c	35	27.2	1399	US-08-471-033-24	Sequence 24, Appl1
c	36	27.2	1399	US-08-471-044-24	Sequence 24, Appl1
c	37	27.2	1399	US-08-463-483A-24	Sequence 24, Appl1
c	38	27.2	1399	US-08-471-046A-24	Sequence 24, Appl1
c	39	27.2	1399	US-08-470-566B-24	Sequence 24, Appl1
c	40	27.2	26.9	US-08-469-334-24	Sequence 24, Appl1
c	41	27.2	1399	US-09-300-529-24	Sequence 24, Appl1
c	42	27	35081	US-08-752-760A-1	Sequence 1, Appl1
c	43	27	50937	US-09-428-517-1	Sequence 1, Appl1
c	44	26.6	900	PCT-US95-04801-3	Sequence 3, Appl1
c	45	26.6	3181	US-08-655-086-1	Sequence 1, Appl1

ALIGNMENTS

```
RESULT 1
US-09-443-501A-2
; Sequence 2, Application US/09443501A
; Patent No. 6303342
; GENERAL INFORMATION:
; APPLICANT: Kusan Biosciences, Inc.
; APPLICANT: Julien, Bryan
; APPLICANT: Katz, Leonard
; APPLICANT: Khosla, Chaitan
; APPLICANT: Tang, Li
; APPLICANT: Ziemann, Rainer
; TITLE OF INVENTION: Recombinant Methods and Materials for Producing
; TITLE OF INVENTION: Epothilone and Epothilone Derivatives
; FILE REFERENCE: 30062-20031.00
; CURRENT APPLICATION NUMBER: US/09/443,501A
; PRIOR FILING DATE: 1999-11-19
; PRIOR APPLICATION NUMBER: US 60/130,560
; PRIOR FILING DATE: 1999-04-22
; PRIOR APPLICATION NUMBER: US 60/122,620
; PRIOR FILING DATE: 1999-03-03
; PRIOR APPLICATION NUMBER: US 60/119,386
; PRIOR FILING DATE: 1999-02-10
; PRIOR APPLICATION NUMBER: US 60/109,401
; PRIOR FILING DATE: 1998-11-20
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 71989
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic construct
US-09-443-501A-2

Query Match      100.0%; Score 101; DB 4; Length 71989;
Best Local Similarity 100.0%; Pred. No. 2.9e-19;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 AGCTTCGAGGTGCTGCGCCCTCCCGAAGCAAGCACTGAGCTGAGTGGCGGCTACAGCA 60
DB      37570 AGCTTCGAGGTGCTGCGCCCTCCCGAAGCAAGCAAGCACTGAGTGGCGGCTACAGCA 37629

QY      61 GCCTGGAGATGAGCTGCTGATGAGGAGCTAGAGCTCGCAAC 101
DB      37630 GCCTGGAGATGAGCTGCTGATGAGGAGCTAGAGCTCGCAAC 37670

RESULT 2
US-09-335-409-1
; Sequence 1, Application US/09335409
```

```
; Patent No. 6121029
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/335,409
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
; US-09-335-409-1
```

```
Query Match          98.4%; Score 99.4; DB 3; Length 68750;
Best Local Similarity 99.0%; Pred. No. 7.9e-19;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 AGCTTCGACGAGTGTGCGCTCCCGAAGCAAGCTCGACGTGATGCCCGCTCACA 60
DB 43182 AGCTTCGACGAGTGTGCGCTCCCGAAGCAAGCTCGACGTGATGCCCGCTCACA 43241
```

```
QY 61 GCCTGGGAATGACCTGCTGATGGGCTAGAGCTGGCGCAAC 101
DB 43242 GCCTGGGAATGACCTGCTGATGGGCTAGAGCTGGCGCAAC 43282
```

```
RESULT 3
US-09-568-102-1
; Sequence 1, Application US/09568102
; Patent No. 6346404
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,102
; PRIOR FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
; US-09-568-102-1
```

```
Query Match          98.4%; Score 99.4; DB 4; Length 68750;
Best Local Similarity 99.0%; Pred. No. 7.9e-19;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 AGCTTCGACGAGTGTGCGCTCCCGAAGCAAGCTCGACGTGATGCCCGCTCACA 60
DB 43182 AGCTTCGACGAGTGTGCGCTCCCGAAGCAAGCTCGACGTGATGCCCGCTCACA 43241
```

```
QY 61 GCCTGGGAATGACCTGCTGATGGGCTAGAGCTGGCGCAAC 101
DB 43242 GCCTGGGAATGACCTGCTGATGGGCTAGAGCTGGCGCAAC 43282
```

```
RESULT 4
US-09-567-969-1
```

```
; Sequence 1, Application US/09567969
; Patent No. 6355457
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/567,969
; PRIOR FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
; US-09-567-969-1
```

```
Query Match          98.4%; Score 99.4; DB 4; Length 68750;
Best Local Similarity 99.0%; Pred. No. 7.9e-19;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 AGCTTCGACGAGTGTGCGCTCCCGAAGCAAGCTCGACGTGATGCCCGCTCACA 60
DB 43182 AGCTTCGACGAGTGTGCGCTCCCGAAGCAAGCTCGACGTGATGCCCGCTCACA 43241
```

```
QY 61 GCCTGGGAATGACCTGCTGATGGGCTAGAGCTGGCGCAAC 101
DB 43242 GCCTGGGAATGACCTGCTGATGGGCTAGAGCTGGCGCAAC 43282
```

```
RESULT 5
US-09-568-480-1
; Sequence 1, Application US/09568480
; Patent No. 6355458
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,480
; PRIOR FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
; US-09-568-480-1
```

```
Query Match          98.4%; Score 99.4; DB 4; Length 68750;
Best Local Similarity 99.0%; Pred. No. 7.9e-19;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 AGCTTCGACGAGTGTGCGCTCCCGAAGCAAGCTCGACGTGATGCCCGCTCACA 60
DB 43182 AGCTTCGACGAGTGTGCGCTCCCGAAGCAAGCTCGACGTGATGCCCGCTCACA 43241
```

```
QY 61 GCCTGGGAATGACCTGCTGATGGGCTAGAGCTGGCGCAAC 101
DB 43242 GCCTGGGAATGACCTGCTGATGGGCTAGAGCTGGCGCAAC 43282
```

```

RESULT - 6
US-09-568-486-1
: Sequence 1, Application US/09568486
: Patent No. 6355459
: GENERAL INFORMATION:
: APPLICANT: Schupp, Thomas
: APPLICANT: Ligon, James
: APPLICANT: Molnar, Istvan
: APPLICANT: Zirkle, Ross
: APPLICANT: Cyr, Devon
: APPLICANT: Goerlach, Joern
: TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
: FILE REFERENCE: 4-30582A
: CURRENT APPLICATION NUMBER: US/09/568,486
: CURRENT FILING DATE: 2000-05-10
: PRIOR APPLICATION NUMBER: 09/335,409
: PRIOR FILING DATE: 1999-06-17
: NUMBER OF SEQ ID NOS: 30
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 1
: LENGTH: 68750
: TYPE: DNA
: ORGANISM: Sorangium cellulosum
: US-09-568-486-1

```

	Query Match	Best Local Similarity	98.4%	Score 99.4%	DB 4	Length 68750
	Matches 100%	Conservative	99.0%	Pred. No. 7,9e19	Mismatches 1	Indels 0
					Gaps 0	
OY	1	AGCTTCGCAGGTGCTGCTGGCCCTCCCGGAAAGCAACTGACGTGATGTCGGCCGCTCACA	60			
Db	43182	AGGTTCTCCAGAGTGTGTGGCGCCCTCTCCGAAGGCAACTGCACGTGATGTCGGCCGCTCACA	43241			
OY	61	GCTTGGGAATGACTGCTGATGGGGGCTAGACTGCGCAAC	101			
OY	43242	GCTTGGGAATGACTGCTGATGGGGGCTAGACTGCGCAAC	43282			

```

RESULT 7
US-09-568-472-1
: Sequence 1, Application US/09568472
: Patent No. 6358719
: GENERAL INFORMATION:
: APPLICANT: Schupp, Thomas
: APPLICANT: Ligon, James
: APPLICANT: Molnar, Istvan
: APPLICANT: Zirkle, Ross
: APPLICANT: Cyf, Devon
: APPLICANT: Goerlach, Joern
: TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
: FILE REFERENCE: 4-30582A
: CURRENT APPLICATION NUMBER: US/09/568,472
: CURRENT FILING DATE: 2000-05-10
: PRIOR APPLICATION NUMBER: 09/335,409
: PRIOR FILING DATE: 1999-06-17
: NUMBER OF SEQ ID NOS: 30
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 1
: LENGTH: 68750
: TYPE: DNA
: ORGANISM: Sorangium cellulosum
US-09-568-472-1

```

Query Match	98.4%	Score 99.4	DB 4	Length 68750
Best Local Similarity	99.0%	Pred. No. 7.9e-19		
Matches 100	Conservative 0	Mismatches 1	Indels 0	Gaps 0
OY	1	AGGTCGCCAGAGTGTGGCGCTCCCGGAAGCAGACTGACGTGGATGGCCGCTCAGCA	60	
Db	43182	AGGTCGCCAGAGTGTGGCGCTCTCCGGAAGCAGACTGACGTGGATGGCCGCTCAGCA	43241	
OY	61	GCCGTGGAGTGGACTCGCTGATGGGGCTAGAGCTCGGCAAC	101	

Db 43242 GCCTGGGATGGACTCGTGATGGGGCTAGAGCTGGCGAAC 43282

```

RESULT 8
US-09-567-899-1
: Sequence 1, Application US/09567899
: Patent No. 6383787
: GENERAL INFORMATION:
: APPLICANT: Schupp, Thomas
: APPLICANT: Lidgon, James
: APPLICANT: Molnar, Istvan
: APPLICANT: Zirkle, Ross
: APPLICANT: Cyr, Devon
: APPLICANT: Goerlach, Joern
: TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
: FILE REFERENCE: 4-30582A
: CURRENT APPLICATION NUMBER: US/09/567,899
: CURRENT FILING DATE: 2000-05-10
: PRIOR APPLICATION NUMBER: 09/335,409
: PRIOR FILING DATE: 1999-06-17
: NUMBER OF SEQ ID NOS: 30
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 1
: LENGTH: 68750
: TYPE: DNA
: ORGANISM: Sorangium cellulosum
: US-09-567-899-1

```

	Query Match	98.4%	Score 99.4	DB 4	Length 68750
	Best Local Similarity	99.0%	Pred. No. 7.9e-19		
	Matches 100	Conservative	0	Mismatches 1	Indels 0
OY	1 AGGTGTCAGAGTGGTGCGCCGCCGAAGAACAATGCATGGATGGCGCTACAGA	60			
Db	43182 AAGTTCACAGGTGGTGCGCCCTCCTCGAAGCAACTGCATGGATGGCGCTACAGA	43241			
OY	61 GCTTGGAAATGACATCGCTGATGGGGCTAGAGACTCGCAAC	101			
Db	43242 GCCTGGAAATGACATCGCTGATGGGGCTAGAGACTCGCAAC	43282			

```

RESULT 9
US-09-144-085-3
: Sequence 3, Application US/09144085
: Patent No. 6280999
: GENERAL INFORMATION:
: APPLICANT: Gustafsson, Claes
: APPLICANT: Betlach, Mary C.
: APPLICANT: Ashley, Gary
: APPLICANT: Ullien, Bryan
: APPLICANT: Ziemann, Rainer
: TITLE OF INVENTION: SORANGIUM POLYKETIDE SYNTHASES AND ENCODING DNA
: FILE OF INVENTION: THEREFOR
: FILE REFERENCE: 30062-20020..20
: CURRENT APPLICATION NUMBER: US/09/144,085
: CURRENT FILING DATE: 1998-08-31
: EARLIER APPLICATION NUMBER: 09/010,809
: EARLIER FILING DATE: 1998-01-22
: NUMBER OF SEQ ID NOS: 8
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 3
: LENGTH: 33529
: TYPE: DNA
: ORGANISM: Sorangium cellulosum
: US-09-144-085-3

```

	Query Match	34.9%	Score 35.2;	DB 4;	Length 33529;
	Best Local Similarity	65.4%	Pred. No. 0.22;		
	Matches 68; Conservative	0;	Mismatches 33;	Indels 3;	Gaps 1;
OY	1 AGGTCCTCCAGTAGTGTGGCCTCCCCGGAAGCAACTGCACGTGCATCGCGC---CGCTCA	57			

Db 4830 AGGTGCGCCGCGGTCTCTGCGGCTCCAGCGCAGCGGCCCTGGGGGGCGAGCAGGTGCTGA 4889

Qy 58 CGAGCCTGGGATGACTCGCTGATGGGGCTAGAGCTGCGCAAC 101
+ ||||| + ||||| ||||| + ||||| |||||
Db 4890 AGGGCCTCGGGCTGACTCGCTATATGGCGGTGAGAGCTGCGCAAC 4933

```

US-08-804-227C-1
US RESULT 10
Sequence 1 Application US/08804227C
Patent No. 5876991
GENERAL INFORMATION:
APPLICANT: Dehoff, Bradley S.
APPLICANT: Kubloss, Stuart A.
APPLICANT: Rosteck, Paul R., Jr.
APPLICANT: Sutton, Kimberly L.
TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: THOMAS G. PLANT 1501
STREET: LILLY CORPORATE CENTER
CITY: INDIANAPOLIS
STATE: IN
COUNTRY: USA
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: ASCII(DOS) Text only
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/804,227C
FILING DATE: February 21, 1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Plant, Thomas, G.
REGISTRATION NUMBER: 35,784
REFERENCE/DOCKET NUMBER: X-8231
TELECOMMUNICATION INFORMATION:
TELEPHONE: 317-276-2439
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 43280 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: Linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 816..14234
FEATURE:
NAME/KEY: CDS
LOCATION: 14351..19945
FEATURE:
NAME/KEY: CDS
LOCATION: 20010..31199
FEATURE:
NAME/KEY: CDS
LOCATION: 31232..36067
FEATURE:
NAME/KEY: CDS
LOCATION: 36249..41774
US-08-804-227C-1

```

Query Match	32.5%;	Score 32.8;	DB 2;	Length 43280;
Best Local Similarity	58.0%;	Pred. No. 0.99;		
Matches 58;	Conservative 0;	Mismatches 42;	Indels 0;	Gaps 0;

[illegible]

Db 30770 CCTCGGCTCGACTCGCTACCGCCGTGAGATGCCAAC 30809

```

RESULT 11
US-07-642-734C-3
Sequence 3: Application US/07642734C
Patent No. 5824513
GENERAL INFORMATION:
APPLICANT: Katz, L
APPLICANT: Donadio, S
APPLICANT: Mcalpine, J B
TITLE OF INVENTION: Recombinant DNA Method for Producing
TITLE OF INVENTION: Erythromycin Analogs
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: Edward H. Gorman
STREET: Abbott Laboratories D377/APed-2 One Abbott
STREET: Park Rd
City: Abbott Park
STATE: IL
COUNTRY: US
ZIP: 60064-3500
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/642,734C
FILING DATE: 17-JAN-91
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Danckers, Andreas M
REGISTRATION NUMBER: 32652
REFERENCE/DOCKET NUMBER: 4952.US.01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 708-937-9396
TELEFAX: 708-938-2623
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 20235 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Saccharopolyspora erythraea
STRAIN: NRRL 238
FEATURE:
NAME/KEY: CDS
LOCATION: 19..10722
OTHER INFORMATION: /codon_start= 19
OTHER INFORMATION: /function= "gene eryA"
OTHER INFORMATION: /product= "eryA ORF2 encoding modules 3 & 4 for
OTHER INFORMATION: 6-deoxyerythronolide B"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 19..4470
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 97..1482
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoacyl ACP synthase of module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1693..2670
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyltransferase domain module 3"

```


FEATURE:
NAME/KEY: misc_feature
LOCATION: 3406..3921
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: beta-ketoreductase domain of module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 4171..4428
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: acyl carrier domain of module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 4471..10722
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: module 4"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 4471..5847
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: beta-ketoacylACPSynthase domain of module"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 6054..7026
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: acyltransferase domain of module 4"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 7165..9216
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: dehydratase and enoylreductase domains m"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 9433..9984
OTHER INFORMATION: /function="approximate span
OTHER INFORMATION: beta-ketoreductase of module 4"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 10225..10483
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: acyl carrier domain of module 4"
FEATURE:
NAME/KEY: CDS
LOCATION: 10723..20235
OTHER INFORMATION: /codon_start= 10723
OTHER INFORMATION: /function="gene-eryA"
OTHER INFORMATION: /product="orf3 encoding modules 5 & 6
OTHER INFORMATION: 6-deoxyerythrinolide B formatio"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 10723..15165
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: module 5"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 10831..12174
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: beta-ketoacylACPSynthase domain of modul"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 12379..13350
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: acyltransferase domain of module 5"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 14062..14610
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: beta-ketoreductase of module 5"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 14857..15114
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: acyl carrier domain of module 5"
FEATURE:

NAME/KEY: misc_feature
LOCATION: 15166..20235
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: module 6"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 15172..16569
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: beta-ketoacylACPSynthase domain of modul"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 16768..17721
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: acyltransferase domain of module 6"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 18379..18921
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: beta-ketoreductase domain of module 6"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 19149..19398
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: acyl carrier domain of module 6"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 19492..20235
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: thioesterase domain of module 6"
US-07-642-734C-3

Query Match 32.3%; Score 32.6; DB 1; Length 20235;
Best Local Similarity 63.3%; Pred. No. 1.1;
Matches 50; Conservative 0; Mismatches 29; Indels 0; Gaps 0;
QY 23 CCCGAAGCGACCTGACGTGATGCGCCGCTCAGAGCCCTGGATGAGCTCGCTGAT 82
DB 10308 CTCGGCGGACGACCTGCCGAGCGCAAGCGCTTCAGAGACCTCGTTGACCTCGCTGCC 10367
QY 83 GGGGCTAGAGCTCGGCAC 101
DB 10368 CGCGGTGAGACTCGGCAC 10386

RESULT 12
US-08-439-009A-3
Sequence 3, Application US/08439009A
Patent No. 6004787
GENERAL INFORMATION:
APPLICANT: Donadio, S
APPLICANT: Meaizine, J B
TITLE OF INVENTION: Method of Directing Biosynthesis of
NUMBER OF INVENTIONS: Specific Polypeptides
CORRESPONDENCE ADDRESS:
ADDRESSER: Steven F. Weinstein
STREET: Abbott Laboratories D377/AD6D-2 One Abbott
CITY: Abbott Park
STATE: IL
COUNTRY: US
ZIP: 60064-3500
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/439,009A
FILING DATE: 11-MAY-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:

NAME: Casuto, Dianne
REGISTRATION NUMBER: 40,943
REFERENCE/DOCKET NUMBER: 4952-US.D1
TELEPHONE: 847-938-3137
TELECOMMUNICATION INFORMATION:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 20235 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
STRAIN: NRRL 238
FEATURE:
NAME/KEY: CDS
LOCATION: 19..10722
OTHER INFORMATION: /codon_start= 19
OTHER INFORMATION: /function= "gene eryA"
OTHER INFORMATION: /product= "eryA ORF2 encoding modules 3 & 4 for
OTHER INFORMATION: 6-deoxyerythronolide B"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 19..4470
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 97..1482
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoacyl ACP synthase of module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1693..2670
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyltransferase domain module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 3406..3921
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoreductase domain of module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 4171..4428
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyl carrier domain of module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 4471..10722
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: module 4"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 4471..5847
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoacylCpsynthase domain of module"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 6054..7026
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyltransferase domain of module 4"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 7165..9216
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: dehydratase and enoylreductase domains m"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 9433..9984
OTHER INFORMATION: /function= "approximate span

OTHER INFORMATION: beta-ketoreductase of module 4"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 10225..10483
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyl carrier domain of module 4"
FEATURE:
NAME/KEY: CDS
LOCATION: 10723..20235
OTHER INFORMATION: /codon_start= 10723
OTHER INFORMATION: /function= "gene -eryA"
OTHER INFORMATION: /product= "orf3 encoding modules 5 & 6
OTHER INFORMATION: 6-deoxyerythronolide B formatio"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 10723..15165
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: module 5"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 10831..12174
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoacylCpsynthase domain of modul"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 12379..13350
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyltransferase domain of module 5"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 14062..14610
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoreductase of module 5"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 14657..15114
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyl carrier domain of module 5"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 15166..20235
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: module 6"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 15172..16569
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoacylCpsynthase domain of modul"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 16768..17721
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyltransferase domain of module 6"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 18379..18921
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoreductase domain of module 6"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 19149..19398
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyl carrier domain of module 6"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 19492..20235
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: thioesterase domain of module 6"
US-08-439-009A-3

Query Match 32.3%; Score 32.6; DB 3; Length 20235;
Best Local Similarity 63.3%; Pred. No. 1.1;
Matches 50; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 6, 2002, 11:51:49 : Search time 157 Seconds
(without alignments)
1448.738 Million cell updates/sec

Title: US-09-724-876-2_COPY_37570_37670

Perfect score: 101
Sequence: 1 aggtctgcagtgctgctgcgc.....tggggctagagctgcgcac 101

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 segs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
N.Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	101	100.0	71989	21	AAA29349
2	99.4	98.4	68750	21	AAZ55887
3	40.2	39.8	5292	22	AAF90036
4	40.2	39.8	34071	22	AAF90033
5	40.2	39.8	42717	22	AAF90032
6	35.2	34.9	33529	23	AAAF17367
7	34.2	33.9	4615	22	AAF90034
8	33	32.7	6564	22	AAF31637
9	32.8	32.5	43280	18	AA780413

10	32.6	32.3	29879	14	AAO46806	eryA region of S.
11	32.2	31.9	53789	19	AAV21187	Amycolatopsis medi
12	31.6	31.3	77536	21	AAA14651	Nucleotide sequenc
13	31.4	31.1	4466	21	AAA14663	Nucleotide sequenc
14	31.4	31.1	4478	21	AAA14661	Nucleotide sequenc
15	31.4	31.1	4547	21	AAA14664	Nucleotide sequenc
16	31.4	31.1	4571	21	AAA14662	Nucleotide sequenc
17	31.4	31.1	77536	21	AAA14651	Nucleotide sequenc
18	31.2	30.9	6459	22	AAF88336	S. spinosa DNA fra
19	31.2	30.9	7788	22	AAF88335	S. spinosa DNA fra
20	31.2	30.9	50000	22	AAF88312	S. spinosa DNA fra
21	31.2	30.9	50000	22	AAF88313	S. spinosa DNA fra
22	31.2	30.9	50000	22	AAF88316	S. spinosa DNA fra
23	31.2	30.9	80161	20	AAZ21501	DNA fragment of Sa
24	31.2	30.9	125401	22	AAZ17186	Streptomyces nours
25	31	30.7	47981	22	AAF30757	Micromonospora meg
26	30.8	30.5	14806	22	AAD03809	Streptomyces gal11
27	30.6	30.3	1419	21	AAC55827	S. lavendulae Mmcm
28	30.6	30.3	6462	22	AAF90037	Nucleotide sequenc
29	30.6	30.3	53500	21	AAC55842	Complete nucleotid
30	30.4	30.1	2237	24	ABL62699	Colon adenocarcino
31	30.4	30.1	3209	21	AACT78080	Human cancer assoc
32	30.4	30.1	5676	19	AAV21186	Amycolatopsis medi
33	30.4	30.1	8460	18	AA788206	cDNA for protein (
34	30.4	30.1	8470	21	AAZ95007	Cancer specific ge
35	30.4	30.1	8519	21	AAA37760	Human fatty acid s
36	30.4	30.1	44377	18	AA78508	Platanolide syntha
37	30.4	30.1	44377	18	AA780414	Platanolide syntha
38	29.8	29.5	11220	21	AAZ87298	S. venezuelae macr
39	29.8	29.5	14071	23	ABL50560	Micromonospora car
40	29.8	29.5	36778	21	AAZ87318	S. venezuelae pik
41	29.8	29.5	37948	21	AAZ87285	Nucleotide sequenc
42	29.8	29.5	38506	21	AAZ75633	Recombinant cosmid
43	29.8	29.5	38506	21	AAZ56001	Mycobacterium tube
44	29.8	29.5	4403765	22	AA199663	Mycobacterium tube
45	29.8	29.5	4411529	22	AA199682	Mycobacterium tube

ALIGNMENTS

RESULT 1	AAA29349	standard; DNA; 71989 BP.
AC	AAA29349;	
XX	12-SEP-2000 (first entry)	
DE	Sorangium cellulosum epothilone polyketide synthase operon genomic DNA.	
XX		
KW	Epothilone; polyketide synthase; epoA; epoB; epoC; epoD; epoE; epoF;	
KW	epoL; epoK; P450 epoxidase; ORFA; ORB; promoter; enhancer; anti-fungal;	
XX	tubulin polymerization assay; anti-tumour; cytostatic; ds.	
XX		
OS	Sorangium cellulosum.	
XX		
Key	Location/Qualifiers	
FT	3..992	
FT	/*tag= a	
FT	/label= ORF_A	
FT	/product= transposase	
FT	/note= "not part of the PKS"	
FT	989..1501	
FT	/*tag= b	
FT	/label= ORF_B	
FT	/product= transposase	
FT	/note= "not part of the PKS"	
FT	1998..6263	
FT	/*tag= c	
FT	/label= epoA_gene	
FT	/note= "encodes the loading domain"	
FT	2031..3548	
FT	misc_RNA	

```
FT      /tag- d
FT      /note- "encodes ketide synthase (KS-Y) of the loading
FT      domain"
FT      misc_RNA
FT      3621..4661
FT      /tag- e
FT      /note- "encodes acyl transferase (AT) of the loading
FT      domain"
FT      misc_RNA
FT      4917..5810
FT      /tag- f
FT      /note- "encodes enoyl reductase (ER) of the loading
FT      domain, potentially involved in formation of the
FT      thiazole moiety"
FT      misc_RNA
FT      5856..6155
FT      /tag- g
FT      /note- "encodes acyl carrier protein (ACP) of the loading
FT      domain"
FT      CDS
FT      6260..10493
FT      /tag- h
FT      /label- epob_gene
FT      /note- "encodes module 1, the NRPS module"
FT      misc_RNA
FT      2031..3548
FT      /tag- i
FT      /note- "encodes condensation domain C2 of the NRPS
FT      module"
FT      misc_RNA
FT      2031..3548
FT      /tag- j
FT      /note- "encodes condensation domain C2 of the NRPS
FT      module"
FT      misc_RNA
FT      6861..6887
FT      /tag- k
FT      /note- "encodes heterocyclization signature sequence"
FT      misc_RNA
FT      6861..6887
FT      /tag- l
FT      /note- "encodes condensation domain C4 of the NRPS
FT      module"
FT      misc_RNA
FT      7358..7366
FT      /tag- m
FT      /note- "encodes condensation domain C7 (partial) of the
FT      NRPS module"
FT      misc_RNA
FT      7898..7921
FT      /tag- n
FT      /note- "encodes adenylation domain A1 of the NRPS module"
FT      misc_RNA
FT      7898..7921
FT      /tag- o
FT      /note- "encodes adenylation domain A1 of the NRPS module"
FT      misc_RNA
FT      8261..8308
FT      /tag- p
FT      /note- "encodes adenylation domain A3 of the NRPS module"
FT      misc_RNA
FT      8411..8422
FT      /tag- q
FT      /note- "encodes adenylation domain A4 of the NRPS module"
FT      misc_RNA
FT      8861..8905
FT      /tag- r
FT      /note- "encodes adenylation domain A6 of the NRPS module"
FT      misc_RNA
FT      8966..8983
FT      /tag- s
FT      /note- "encodes adenylation domain A7 of the NRPS module"
FT      misc_RNA
FT      9090..9179
FT      /tag- t
FT      /note- "encodes adenylation domain A8 of the NRPS module"
FT      misc_RNA
FT      9183..9992
FT      /tag- u
FT      /note- "encodes oxidation region for forming thiazole"
FT      misc_RNA
FT      10121..10138
FT      /tag- v
FT      /note- "encodes adenylation domain A10 of the NRPS
FT      module"
FT      misc_RNA
FT      10261..10306
FT      /tag- w
FT      /note- "encodes thiolation domain (PCP) of the NRPS
FT      module"
FT      CDS
FT      10639..16137
FT      /tag- x

FT      /label- epoc_gene
FT      /note- "encodes module 2"
FT      misc_RNA
FT      10654..12033
FT      /tag- y
FT      /note- "encodes KS2, the KS domain of module 2"
FT      misc_RNA
FT      12250..13287
FT      /tag- z
FT      /note- "encodes AT2, the AT domain of module 2"
FT      misc_RNA
FT      13327..13899
FT      /tag- aa
FT      /note- "encodes dehydratase (DH) 2, the DH domain of
FT      module 2"
FT      misc_RNA
FT      14962..15756
FT      /tag- ab
FT      /note- "encodes ketoreductase (KR) 2, the KR domain of
FT      module 2"
FT      misc_RNA
FT      15763..16008
FT      /tag- ac
FT      /note- "encodes ACP2, the ACP domain of module 2"
FT      CDS
FT      16134..37907
FT      /tag- ad
FT      /label- epod_gene
FT      /note- "encodes modules 3-6"
FT      misc_RNA
FT      16425..17606
FT      /tag- ae
FT      /note- "encodes KS3"
FT      misc_RNA
FT      17817..18857
FT      /tag- af
FT      /note- "encodes AT3"
FT      misc_RNA
FT      19581..20396
FT      /tag- ag
FT      /note- "encodes KR3"
FT      misc_RNA
FT      20424..20642
FT      /tag- ah
FT      /note- "encodes ACP3"
FT      misc_RNA
FT      20706..22082
FT      /tag- ai
FT      /note- "encodes KS4"
FT      misc_RNA
FT      22296..23336
FT      /tag- aj
FT      /note- "encodes AT4"
FT      misc_RNA
FT      24069..24647
FT      /tag- ak
FT      /note- "encodes KR4"
FT      misc_RNA
FT      24867..25151
FT      /tag- al
FT      /note- "encodes ACP4"
FT      misc_RNA
FT      25203..26576
FT      /tag- am
FT      /note- "encodes KS5"
FT      misc_RNA
FT      26793..27883
FT      /tag- an
FT      /note- "encodes AT5"
FT      misc_RNA
FT      27966..28574
FT      /tag- ao
FT      /note- "encodes DH5"
FT      misc_RNA
FT      29433..30287
FT      /tag- ap
FT      /note- "encodes ER5"
FT      misc_RNA
FT      30321..30869
FT      /tag- aq
FT      /note- "encodes KR5"
FT      misc_RNA
FT      31077..31373
FT      /tag- ar
FT      /note- "encodes ACP5"
FT      misc_RNA
FT      31440..32807
FT      /tag- as
FT      /note- "encodes KS6"
FT      misc_RNA
FT      33018..34067
FT      /tag- at
FT      /note- "encodes AT6"
FT      misc_RNA
FT      34107..34676
FT      /tag- au
```

```
FT /note= "encodes DH6"
FT misc_RNA 35760..36641
FT /tag= av
FT /note= "encodes ER6"
FT misc_RNA 36705..37256
FT /tag= aw
FT /note= "encodes KR6"
FT misc_RNA 37470..37769
FT /tag= ax
FT /note= "encodes ACP6"
FT CDS 37912..49308
FT /tag= ay
FT /label= epof_gene
FT /note= "encodes modules 7 and 8"
FT misc_RNA 38014..39375
FT /tag= az
FT /note= "encodes KS7"
FT misc_RNA 39589..40626
FT /tag= ba
FT /note= "encodes AT7"
FT misc_RNA 41341..41922
FT /tag= bb
FT /note= "encodes KR7"
FT misc_RNA 42181..42423

Query Match 100.0%; Score 101; DB 21; Length 71989;
Best Local Similarity 100.0%; Pred. No. 1.2e-19;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGCTCTCGCAGGTGCTGGCCCTCCCGAAGCAAGCTGAGCTGGATCGCCGCTCACA GA 60
Db 37570 AGGTTCTCGCAGGTGCTGGCCCTCCCGAAGCAAGCTGAGCTGGATCGCCGCTCACA GA 37629
Qy 61 GCCTGGGAATGAGCTCGCTGATGGGGCTAGAGCTCGCAAC 101
Db 37630 GCCTGGGAATGAGCTCGCTGATGGGGCTAGAGCTCGCAAC 37670

RESULT 2
ID AA255887 standard; DNA; 68750 BP.
AC AA255887;
XX
XX 10-APR-2000 (first entry)
DE Sorangium cellulosum 68.75 kb contig.
XX
XX Epithione biosynthesis; type I polyketide synthase; taxol substitute;
KW anticancer; ds.
OS Sorangium cellulosum.
XX
XX Location/Qualifiers
FH 1..1826
FT CDS
FT /tag= a
FT /partial
FT /product= "Partial Orf 1 protein (AAV58580)"
FT /note= "No initiation codon given in the specification"
FT CDS
FT /tag= b
FT /product= "Orf 2 protein (AAV58581)"
FT 3415..5556
FT /tag= c
FT /product= "Orf 3 protein (AAV58582)"
FT CDS
FT /tag= d
FT /product= "Orf 4 protein (AAV58583)"
FT 6226..6675
FT /tag= e
FT /product= "Orf 5 protein (AAV58584)"
FT 7610..11875
FT /tag= f

FT /product= "Type I polyketide synthase, EPOS A
FT (AAV58573)"
FT 11872..116104
FT /tag= g
FT /product= "Non-ribosomal peptide synthetase, EPOS P
FT (AAV58574)"
FT 16251..21749
FT /tag= h
FT /product= "Type I polyketide synthase, EPOS B
FT (AAV58575)"
FT 21746..43519
FT /tag= i
FT /product= "Type I polyketide synthase, EPOS C
FT (AAV58576)"
FT 43524..54920
FT /tag= j
FT /product= "Type I polyketide synthase, EPOS D
FT (AAV58577)"
FT 54935..62254
FT /tag= k
FT /product= "Type I polyketide synthase, EPOS E
FT (AAV58578)"
FT 62369..63628
FT /tag= l
FT /product= "Cytochrome P450 oxygenase homologue, EPOS F
FT (AAV58579)"
FT 63779..64333
FT /tag= m
FT /product= "Orf 6 protein (AAV58585)"
FT CDS
FT /product= "Orf 6 protein (AAV58585)"
FT complement (63853..64290)
FT /tag= n
FT /product= "Orf 7 protein (AAV58586)"
FT 64363..64920
FT /tag= o
FT /product= "Orf 8 protein (AAV58587)"
FT CDS
FT /product= "Orf 8 protein (AAV58587)"
FT complement (64287..64727)
FT /tag= p
FT /product= "Orf 9 protein (AAV58588)"
FT 65063..65767
FT /tag= q
FT /product= "Orf 10 protein"
FT CDS
FT /product= "Orf 10 protein"
FT complement (65008..65874)
FT /tag= r
FT /product= "Orf 11 protein (AAV58590)"
FT CDS
FT /product= "Orf 11 protein (AAV58590)"
FT complement (65871..66338)
FT /tag= s
FT /product= "Orf 12 protein (AAV58591)"
FT CDS
FT /product= "Orf 12 protein (AAV58591)"
FT 66667..67137
FT /tag= t
FT /product= "Orf 13 protein (AAV58592)"
FT CDS
FT /product= "Orf 13 protein (AAV58592)"
FT 67334..68251
FT /tag= u
FT /product= "Orf 14 protein (AAV58593)"
FT CDS
FT /product= "Orf 14 protein (AAV58593)"
FT 68346..68750
FT /partial
FT /tag= v
FT /product= "Partial Orf 15 protein (AAV58594)"
FT /note= "No termination codon given in the specification"
FT MO9966028-A2.
FT 23-DEC-1999.
FT PD
FT 16-JUN-1999; 99WO-EP04171.
FT XX
FT 18-JUN-1998; 98US-0099504.
FT PR 24-SEP-1998; 98US-0101631.
FT PR 05-FEB-1999; 99US-0118906.
FT XX
FT (NOVS ) NOVARTIS AG.
FT PA (NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.
FT XX Schnupp T, Ligon JM, Molnar I, Zirkle R, Goerlach J, Cyr D;
```

XX WPI: 2000-097741/08.
 DR P-PSDB; AAY58573, AAY58574, AAY58575, AAY58576, AAY58577, AAY58578,
 DR AAY58579, AAY58580, AAY58581, AAY58582, AAY58583, AAY58584,
 DR AAY58585, AAY58586, AAY58587, AAY58588, AAY58590, AAY58591,
 DR AAY58592, AAY58593, AAY58594.
 XX New isolated epothilone synthase genes, used for the recombinant
 PT production of epothilone for use in cancer therapy
 XX
 PS Claim 14; Page 87-104; 174pp; English.
 XX This sequence represents a 68.75 kb contig from *Sorangium cellulosum*
 CC comprising 22 open reading frames (ORFs) and includes genes encoding
 CC proteins involved in the biosynthesis of epothilones. Epothilones A and
 CC B are 16-membered macrocyclic polyketides with an acylcysteine-derived
 CC starter unit; polyketides being synthesised from two-carbon building
 CC blocks, the beta-carbon of which always carries a keto group. Each round
 CC of two-carbon addition is carried out by a complex of enzymes known as
 CC the polyketide synthase in a manner similar to fatty acid biosynthesis.
 CC EPOS A (AAY58573) and EPOS P (AAY58574) are involved in formation of
 CC the thiazole ring formation of epothilones, and EPOS B, EPOS C, EPOS D
 CC and EPOS E (AAY58575-Y58578) are involved in polyketide backbone
 CC formation. EPO F (AAY58579) is an epothilone macrolactone oxidase, and
 CC the proteins Orf 3 (AAY58582) and Orf14 (AAY58593) are thought to be
 CC involved in transport. Epothilones mimic the biological activity of
 CC taxol, and may be substituted for taxol in cancer chemotherapeutic
 CC compositions. Epothilones exhibit a much lower drop in potency against a
 CC multiply drug-resistant cell line compared with taxol, and are
 CC considerably less efficiently exported from such cells by the multidrug
 CC resistance protein (MDR, or P-glycoprotein). Despite the potential of
 CC epothilones as anticancer agents, they are problematical to produce on a
 CC large scale. Epothilones are too complex for industrial scale chemical
 CC synthesis, and *Sorangium cellulosum* is difficult to ferment, producing
 CC poor yields of epothilones. The nucleic acids of the invention may be
 CC used for the recombinant production of epothilones in a heterologous host
 CC that is more amenable to fermentation.
 XX
 SQ Sequence 68750 BP; 9596 A; 22458 C; 25537 G; 11159 T; 0 other;
 XX
 Query Match 98.4%; Score 99.4; DB 21; Length 68750;
 Best Local Similarity 99.0%; Pred. No. 3.5e-19;
 Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 AGGTCGCAAGTGTGCGCCTCCCGAAGCAAGCTGACGTGATGCGCGCTACGA 60
 DB 43182 AGGTCTCGAGGTGTCGCGCTCTCGAAGCAAGCTGACGTGATGCGCGCTACGA 43241
 QY 61 GCCTGGGAATGCATCGCTGATGGGCTAGAGCTGCGCAAC 101
 DB 43242 GCCTGGGAATGCATCGCTGATGGGCTAGAGCTGCGCAAC 43282
 XX
 RESULT 3
 ID AAF90036 standard; DNA; 5292 BP.
 XX AAF90036;
 AC AAF90033;
 XX
 DT 06-AUG-2001 (first entry)
 XX
 DE Nucleotide sequence of a type I polyketide synthase.
 XX
 KW Metabolic pathway operon; polyketide; polyketide antibiotic;
 XX type I polyketide synthase; ss.
 XX
 OS Unidentified.
 XX
 Key Location/Qualifiers
 FH 1..5292
 FT /tag= a
 FT /product= "type I polyketide synthase"
 XX

PN WO200140497-A2.
 XX
 PD 07-JUN-2001.
 XX
 XX 27-NOV-2000; 2000WO-FR03311.
 XX
 XX 29-NOV-1999; 99FR-0015032.
 PR 07-JUN-2000; 2000US-0209800.
 XX
 PA (AVENTIS PHARMA SA.
 XX
 PI Jeanin P, Pernodet J, Guerin M, Simonet P, Courtois S;
 PI Cappellano C, Francou F, Raynal A, Ball M, Sezonov G, Tuphile K;
 PI Frostegard A;
 XX
 DR WPI: 2001-374849/39.
 DR P-PSDB; AAB83973.
 XX
 XX Collection of nucleic acids from environmental samples, useful for
 PT identifying e.g. genes encoding polyketide synthases and derived
 PT antibiotics
 XX
 PS Claim 35; Page 305-307; 356pp; French.
 XX
 CC The specification describes a method for the preparation of a collection
 CC of nucleic acids from organisms in a soil sample. The method comprises
 CC milling a dried sample to produce microparticles; suspending these in
 CC liquid buffer; extraction of nucleic acids from the microparticle;
 CC passing nucleic acid-containing solution through a molecular sieve;
 CC passing nucleic acid-enriched fractions through an anion exchange
 CC chromatography material; and recovering fractions containing purified
 CC nucleic acids. The nucleic acids are sources for sequences that encode
 CC either operons involved in a metabolic pathway (specifically polyketide
 CC synthesis) or polypeptides, particularly for production of therapeutic
 CC or agricultural compounds, especially polyketide antibiotics. AAF90034-39
 CC represent open reading frames (ORFs) of the coding strand of cosmid
 CC a26g1, and encode type I polyketide synthases.
 XX
 SQ Sequence 5292 BP; 905 A; 1715 C; 1725 G; 947 T; 0 other;
 XX
 Query Match 39.8%; Score 40.2; DB 22; Length 5292;
 Best Local Similarity 64.5%; Pred. No. 0.022; 33; Indels 0; Gaps 0;
 Matches 60; Conservative 0; Mismatches 33; Indels 0; Gaps 0;
 QY 9 CAGGTCTGCGCTTCCCGAAGCAAGCTGACGTGATGCGCGCTACGAAGCTGCGGA 68
 DB 4885 CAGGTCTGCGCTTCCCGAAGCAAGCTGACGTGATGCGCGCTACGAAGCTGCGGA 4944
 QY 69 ATGGAATCGCTGATGGGCTAGAGCTGCGCAAC 101
 DB 4945 TTGCAATCGTGTGATGGGCTCGCGCAAGTTCGCAAC 4977
 XX
 RESULT 4
 ID AAF90033 standard; DNA; 34071 BP.
 XX AAF90033;
 AC AAF90033;
 XX
 DT 06-AUG-2001 (first entry)
 XX
 DE Nucleotide sequence of cosmid a26g1 (coding strand).
 XX
 KW Metabolic pathway operon; polyketide; polyketide antibiotic; ss.
 XX
 OS Synthetic.
 XX
 PN WO200140497-A2.
 XX
 PD 07-JUN-2001.
 XX
 XX 27-NOV-2000; 2000WO-FR03311.
 XX

PR 29-NOV-1999; 99FR-0015032.
PR 07-JUN-2000; 2000US-0209800.
XX
XX (AVET) AVENTIS PHARMA SA.
XX
PI Jeannin P, Pernodet J, Guerlneau M, Simonet P, Courtois S;
PI Cappellano C, Francou F, Raynal A, Ball M, Sezonov G, Tuphile K;
PI Frostegard A;
XX
XX WPI: 2001-374849/39.
DR
XX
XX Collection of nucleic acids from environmental samples, useful for
PT identifying e.g. genes encoding polyketide synthases and derived
PT antibiotics -
XX
PS Example 14; Page 289-300; 356pp; French.
XX
XX The specification describes a method for the preparation of a collection
CC of nucleic acids from organisms in a soil sample. The method comprises
CC milling a dried sample to produce microparticles; suspending these in
CC liquid buffer; extraction of nucleic acids from the microparticle;
CC passing nucleic acid-containing solution through a molecular sieve;
CC chromatography material; and recovering fractions containing purified
CC nucleic acids. The nucleic acids are sources for sequences that encode
CC either operons involved in a metabolic pathway (specifically polyketide
CC synthesis) or polypeptides, particularly for production of therapeutic
CC or agricultural compounds, especially polyketide antibiotics. The present
CC sequence represents cosmid a2691 coding strand), which encodes different
CC polyketide synthases.
XX
SQ Sequence 34071 BP; 5791 A; 10858 C; 11089 G; 6333 T; 0 other;
XX
Query Match 39.8%; Score 40.2; DB 22; Length 34071;
Best Local Similarity 64.5%; Pred. No. 0.026;
Matches 60; Conservative 0; Mismatches 33; Indels 0; Gaps 0;
XX
OY 9 CAGGTCGTGGCGCTCCCGAAGCGCAAGCTCGACGTGGCGCTCAGAGCCTGGGA 68
DB 17820 CACGTGCTGCGCTTGTGATCCAGCAACTCGACGACGACGCTGGTGTCTCGGA 17879
|| ||||| ||||| || || ||||| ||||| ||||| |||||
OY 69 ATGAGCTCGCTGATGGGCTAGAGCTGCGCAAC 101
|| ||||| ||||| || || ||||| ||||| |||||
DB 17880 TTCGATTCGTGATGGCCCTCGAGTTTCGAAC 17912
XX
RESULT 5
AAF90032/C
ID AAF90032 standard; DNA; 42717 BP.
XX
XX AAF90032;
AC
XX 06-AUG-2001 (first entry)
DT
XX
XX Nucleotide sequence of cosmid a2691 (non-coding strand).
DE
XX Metabolic pathway operon; polyketide; polyketide antibiotic; ss.
KW
XX Synthetic.
OS
XX W0200140497-A2.
PN
XX 07-JUN-2001.
PD
XX
XX 27-NOV-2000; 2000WO-FR03311.
PF
XX
XX 29-NOV-1999; 99FR-0015032.
PR
XX 07-JUN-2000; 2000US-0209800.
XX
XX (AVET) AVENTIS PHARMA SA.
PA
XX
XX Jeannin P, Pernodet J, Guerlneau M, Simonet P, Courtois S;
PI
PI Cappellano C, Francou F, Raynal A, Ball M, Sezonov G, Tuphile K;

PI Frostegard A;
XX
XX WPI: 2001-374849/39.
DR
XX
XX Collection of nucleic acids from environmental samples, useful for
PT identifying e.g. genes encoding polyketide synthases and derived
PT antibiotics -
XX
PS Example 14; Page 274-288; 356pp; French.
XX
XX The specification describes a method for the preparation of a collection
CC of nucleic acids from organisms in a soil sample. The method comprises
CC milling a dried sample to produce microparticles; suspending these in
CC liquid buffer; extraction of nucleic acids from the microparticle;
CC passing nucleic acid-containing solution through a molecular sieve;
CC passing nucleic acid-enriched fractions through an anion exchange
CC chromatography material; and recovering fractions containing purified
CC nucleic acids. The nucleic acids are sources for sequences that encode
CC either operons involved in a metabolic pathway (specifically polyketide
CC synthesis) or polypeptides, particularly for production of therapeutic
CC or agricultural compounds, especially polyketide antibiotics. The present
CC sequence represents cosmid a2691 (non-coding strand). The sense strand
CC encodes different polyketide synthases.
XX
SQ Sequence 42717 BP; 8230 A; 13520 C; 13184 G; 7782 T; 1 other;
XX
Query Match 39.8%; Score 40.2; DB 22; Length 42717;
Best Local Similarity 64.5%; Pred. No. 0.027;
Matches 60; Conservative 0; Mismatches 33; Indels 0; Gaps 0;
XX
OY 9 CAGGTCGTGGCGCTCCCGAAGCGCAAGCTCGACGTGGCGCTCAGAGCCTGGGA 68
DB 16300 CACGTGCTGCGCTTGTGATCCAGCAACTCGACGACGACGCTGGTGTCTCGGA 16241
|| ||||| ||||| || || ||||| ||||| ||||| |||||
OY 69 ATGAGCTCGCTGATGGGCTAGAGCTGCGCAAC 101
|| ||||| ||||| || || ||||| ||||| |||||
DB 16240 TTCGATTCGTGATGGCCCTCGAGTTTCGAAC 16208
XX
RESULT 6
AAS17367
ID AAS17367 standard; DNA; 33529 BP.
XX
XX AAS17367;
AC
XX 12-MAR-2002 (first entry)
DT
XX
XX DNA sequence of S. cellulosum polyketide synthase cosmid, pK0528-26.
DE
XX
XX Polyketide synthase; PKS; catalytic domain; ketosynthase domain;
KW acyl transferase domain; dehydratase domain; ketoreductase domain;
KW acyl carrier protein domain; pK0528-26; ds.
XX
XX Sorangium cellulosum.
OS
XX
XX US6280999-B1.
PN
XX 28-AUG-2001.
PD
XX
XX 31-AUG-1998; 98US-0144085.
PF
XX
XX 22-JAN-1998; 98US-0010809.
PR
XX
XX (KOSA-) KOSAN BIOSCIENCE.
PA
XX
XX Gustafsson C, Belach MC, Ashley G, Julien B, Ziermann R;
PI
XX WPI: 2001-606536/69.
DR
XX
XX Novel purified, isolated DNA molecule from Sorangium cellulosum having
PT polyketide open reading frame encoding modules with one or more domains
PT such as ketosynthase, acyl transferase and acyl carrier protein domains
PT -

XX Claim 4; Fig 1; 72pp; English.
PS
XX
CC The present invention relates to the isolation of novel Sorangium
CC cellulose polyketide synthases (PKS), and the polynucleotide sequences
CC encoding them. The polyketide synthases include catalytic domains such
CC as ketosynthase domain, acyl transferase domain, dehydratase domain,
CC ketoreductase domain and acyl carrier protein domain. A host cell
CC comprising a PKS ORF (open reading frame) which encodes one or more
CC more PKS domains is useful for producing polyketide synthases from which
CC polyketides can be produced. The host cells are useful for constructing
CC a library, where each individual colony of the library represents a
CC colony with the ability to produce a particular PKS synthase and
CC ultimately a particular polyketide. The polyketides produced by these
CC colonies can be used collectively in a panel to represent a library or
CC may be assessed individually for activity. Colonies in the library are
CC also induced to produce the relevant synthases and thus to produce the
CC relevant polyketides to obtain a library of candidate polyketides which
CC can be screened for binding to desired targets such as receptors,
CC signalling proteins, etc. The present sequence represents the DNA
CC sequence of cosmid pK0528-26 which encodes one or more domains of
CC S. cellulosum PKS.
CC Note: The present sequence is said to encode the functional domains
CC of S. cellulosum PKS which correspond to domains or domain subsets of
CC the amino acid sequences of ORF1 (AAU10700) and ORF2 (AAU10701).
XX
SQ Sequence 33529 BP; 4489 A; 9518 C; 14470 G; 5046 T; 6 other;
Query Match 34.9%; Score 35.2; DB 23; Length 33529;
Best Local Similarity 65.4%; Pred. No. 0.7; Mismatches 33; Indels 3; Gaps 1;
Matches 68; Conservative 0;
QY 1 AGGTCGCGAGTGTCTGCGCCCTCCCGAAGCAAGCTGACGTGATGCGC---CGCTCA 57
DB 4830 AGGTCCGCGGCGTCTGCGGCTCCAGCGACGAGGCCCTGGGGGGCGAGCAGTCTCTGA 4889
QY 58 CGAGCTGGGAATGAGACTGCTGATGGGGCTAAGCTGCCCAAC 101
DB 4890 AGGGCTCTGGGCTGACCTGATGGCGGTGGAGCTGCCCAAC 4933
RESULT 7
AAF90034
ID AAF90034 standard; DNA: 4615 BP.
XX
AC AAF90034;
XX
DT 06-NOV-2001 (first entry)
XX
DE Nucleotide sequence of a type I polyketide synthase.
XX
XX Metabolic pathway operon: polyketide; polyketide antibiotic;
KM type I polyketide synthase; ss.
XX
OS unidentified.
XX
FH Key Location/Qualifiers
FT CDS 2..4615
FT /tag= a
FT /product= "type I polyketide synthase"
XX
PN WO200140497-A2.
XX
PD 07-JUN-2001.
XX
PF 27-NOV-2000; 2000WO-FR03311.
XX
PR 29-NOV-1999; 99FR-0015032.
XX
PR 07-JUN-2000; 2000US-0209800.
XX
PA (AVER) AVENTIS PHARMA SA.
XX
PI Jeanin P, Pernodet J, Guerin M, Simonet P, Courtois S;

PI Cappellano C, Francou F, Raynal A, Bail M, Sezonov G, Tuphile K;
PI Frostegard A;
XX
XX WPI: 2001-374849/39.
DR P-PDB: AAB83971.
XX
XX Collection of nucleic acids from environmental samples, useful for
PT identifying e.g. genes encoding polyketide synthases and derived
PT antibiotics
XX
PS Claim 35; Page 300-302; 356pp; French.
XX
CC The specification describes a method for the preparation of a collection
CC of nucleic acids from organisms in a soil sample. The method comprises
CC milling a dried sample to produce microparticles; suspending these in
CC liquid buffer; extraction of nucleic acids from the microparticle;
CC passing nucleic acid-containing solution through a molecular sieve;
CC passing nucleic acid-enriched fractions through an anion exchange
CC chromatography material; and recovering fractions containing purified
CC nucleic acids. The nucleic acids are sources for sequences that encode
CC either operons involved in a metabolic pathway (specifically polyketide
CC synthesis) or polypeptides, particularly for production of therapeutic
CC or agricultural compounds, especially polyketide antibiotics. AAF90034-39
CC represent open reading frames (ORFs) of the coding strand of cosmid
CC a2691, and encode type I polyketide synthases.
XX
SQ Sequence 4615 BP; 827 A; 1477 C; 1429 G; 882 T; 0 other;
Query Match 33.9%; Score 34.2; DB 22; Length 4615;
Best Local Similarity 62.1%; Pred. No. 1.1; Mismatches 33; Indels 0; Gaps 0;
Matches 54; Conservative 0;
QY 15 CTGCGCTCTCCCGAAGCAAGCTGACGTGATGCGCGCTCAGCAGCCCTGGGAATGAC 74
DB 2 CTGAGTGGCCCGAATGCGGCGGTGACGACTTACAGAGCCGCTGGGTATGGATTGGAC 61
QY 75 TCGCTATGGGCGCTAGAGCTGCCCAAC 101
DB 62 TCGCTATGGGCGGTGCAATTACGCCAAC 88
RESULT 8
AAF31637
ID AAF31637 standard; DNA: 6564 BP.
XX
AC AAF31637;
XX
DT 09-APR-2001 (first entry)
XX
DE Mycobacterium tuberculosis ppsc gene.
XX
XX Mycobacterium tuberculosis; attenuated microorganism;
KM signature tagged transposon mutant; mutant library;
KM mycobacterial infection; actinomycetales; antibacterial; immunostimulant;
KW vaccine; ppsc; ds.
XX
XX Mycobacterium tuberculosis.
OS
OS
PN WO200102555-A1.
XX
PD 11-JAN-2001.
XX
PF 06-JUL-2000; 2000WO-IB00950.
XX
PR 06-JUL-1999; 99US-0142982.
XX
PR 08-JUL-1999; 99US-0142833.
XX
PA (INSP) INSP PASTEUR.
XX
PI Gicquel B, Guilhot C, Camacho L;
XX
DR WPI: 2001-091804/10.
DR P-PDB: AAB66467.

XX Screening a mutant library for mutants unable to grow under specific
PT conditions and for identifying loci involved in pathogenicity,
PT comprises using signature tagged transposon mutagenesis -
XX
PS Example 8; Page 122-124; 159pp; English.
XX
CC The present sequence is given in a specification relating to a method for
CC screening a library of mutants. The method comprises constructing a
CC library with insertions in genes and/or regulatory regions of the
CC organisms of interest, where the insertion contains a tag and/or a
CC transposon associated with a tag. The mutants are identified by
CC hybridisation of the tags to known sequences. The method is useful for
CC treating an individual suffering from a mycobacterial infection,
CC suspected of being infected with a mycobacterium, or having been
CC exposed to an infectious mycobacterium. It is also useful for
CC identifying and isolating mutants of actinomycetales and for identifying
CC compounds that have antibiotic activity. The method is used to identify
CC mutants of microorganisms, preferably an actinomycetales, such as
CC *M. tuberculosis*, *M. bovis*, *M. leprae*, *M. avium*, *M. intracellulare* and
CC *M. paratuberculosis*, that is unable to grow under specific conditions.
CC It is especially useful for identifying loci involved in pathogenicity.
CC It is useful in constructing vaccines. The method can be used to screen
CC multiple libraries concurrently. It can screen libraries of different
CC organisms or different strains of the same organism.
XX
SQ Sequence 6564 BP; 1137 A; 2189 C; 2154 G; 1084 T; 0 other;
Query Match 32.7%; Score 33; DB 22; Length 6564;
Best Local Similarity 60.7%; Pred. No. 2.6;
Matches 54; Conservative 0; Mismatches 35; Indels 0; Gaps 0;
QY 12 GTCGCGCCCTCCCGGAGCGACGCTGATGCGCCCTCAGCAGCTGGGATG 71
Db 6250 GCGGTCTGCTGCTGCGGCGATCCATCGATCACCAGCCGCTGAAACCTGGGACTC 6309
QY 72 GACTCGCTGATGGGCTAGAGCTGGCAA 100
Db 6310 GACTCGCTGATGGGCTGGAATGCGCAA 6338
RESULT 9
AA080413
ID AA080413 standard; DNA: 43280 BP.
XX
AC AA080413;
XX.
DT 27-FEB-1998 (first entry)
XX
DE Tyactone synthase gene cluster.
XX
KW Tyactone synthase gene cluster; tylg gene; multifunctional protein;
KW polyketide; tyactone synthesis; antibiotic; tylosin; ss.
XX
OS Streptomyces fradiae.
XX
FH Key location/Qualifiers
FT 816..14243
FT /*tag= a
FT /transl_except= (pos: 816..818, aa: Met)
FT /note= "ORF1 encodes protein shown in AAW22601"
FT 14351..19945
FT /*tag= b
FT /transl_except= (pos: 14351..14353, aa: Met)
FT /note= "ORF2 encodes protein shown in AAW22602"
FT 20010..31199
FT /*tag= c
FT /transl_except= (pos: 20010..20012, aa: Met)
FT /note= "ORF3 encodes protein shown in AAW22603"
FT 31232..36067
FT /*tag= d
FT /note= "ORF4 encodes protein shown in AAW22604"
FT 36249..41774
FT CDS

FT /*tag= e
FT /note= "ORF5 encodes protein shown in AAW22605"
XX
PN EP791655-A2.
XX
PD 27-AUG-1997.
XX
PF 19-FEB-1997; 97EP-0301056.
XX
PR 22-FEB-1996; 96US-0012078.
XX
PA (ELIT) LILLY & CO ELI.
XX
PI Dehoff BS, Kuhstoss SA, Rosteck PR, Sutton KL;
XX WPI: 1997-418046/39.
XX DR P-PSDB: AAW22601-W22605.
XX
PT DNA encoding Streptomyces fradiae tyactone synthase domain - for
PT production of tylosin-related polyketide compounds
XX
PS Claim 2; Pages 8-66; 220pp; English.
XX
CC This sequence represents the tyactone synthase gene cluster of the
CC invention. This sequence is also referred to as the tylg gene, and was
CC isolated from Streptomyces fradiae. This sequence encodes multifunctional
CC proteins which direct the synthesis of the polyketide tyactone, isolated
CC from Streptomyces fradiae. Tyactone is the basic building block of the
CC antibiotic tylosin. The DNA sequence can be modified so as to alter the
CC type of carboxylic acids incorporated, the number of carboxylic acids
CC incorporated and/or the post-condensation reactions performed, thereby
CC resulting in novel tylosin-related polyketides.
XX
SQ Sequence 43280 BP; 5269 A; 15187 C; 16651 G; 6173 T; 0 other;
Query Match 32.5%; Score 32.8; DB 18; Length 43280;
Best Local Similarity 58.0%; Pred. No. 3.5;
Matches 58; Conservative 0; Mismatches 42; Indels 0; Gaps 0;
QY 2 GGTCTGCGAGTGTGCTGCGCTCCCGAGGCAAGCTCGAGTGGCGCGCTCAGCAG 61
Db 30710 GCGGCGGTGCTGGGCGGCGATCCGACGATCGATCCCGAGGCGGCGCTTCAAGGA 30769
QY 62 CCTGGGAATGAGCTGCTGATGGGCTGAGCTGGGCAAC 101
Db 30770 CCTGGCTTGACTGCTCACCCTGGGATGGGCAAC 30809
RESULT 10
AA046806
ID AA046806 standard; DNA: 29879 BP.
XX
AC AA046806;
XX
DT 22-DEC-1993 (first entry)
XX
DE eryA region of S. erythraea chromosome.
XX
KW Saccaropolyspora erythraea; eryA; biosynthesis; polyketide; module;
KW erythromycin; condensation; elongation; acyl chain growth;
KW gene replacement; ss.
XX
OS Saccaropolyspora erythraea.
XX
FH Key location/Qualifiers
FT 3..10199
FT /*tag= a
FT /note= "ORF 1"
FT 10218..20921
FT /*tag= b
FT /note= "ORF 2"
FT 20922..29879
FT /*tag= c
FT CDS

```

FT XX /note= "ORF 3"
PN MO9313663-A.
XX
XX 22-JUL-1993.
PD
XX 17-JAN-1992; 92MO-US00427.
XX
XX 17-JAN-1992; 92MO-US00427.
XX
XX (ABBO ) ABBOTT LAB.
XX
XX Donadio S, Katz L, McAlpine JB;
XX
XX WPI: 1993-242804/30.
DR P-PSDB; AAR44430-32.
XX
XX Biosynthesis of specific polyketide analogues esp. erythromycin
PT cpds. - by introducing altered biosynthetic gene-contg. DNA into
PT microorganisms
XX
XX Claim 27; Fig 2; 13pp; English.
XX
XX This sequence represents a fragment of the Saccharopolyspora erythraea
CC genome, designated eryA. The polypeptides encoded by this region
CC are involved in the biosynthesis of the polyketide segment of
CC erythromycin. eryA is organised in modules and each module takes care
CC of one condensation step. The precise succession of elongation steps
CC is dictated by the genetic order of the modules. This fragment may be
CC specifically altered such that novel polyketide molecules of desired
CC structure are produced. Three types of alteration may be produced;
CC those inactivating a single function in a module which does not arrest
CC acyl chain growth; those inactivating a single function in a module
CC which does affect chain growth; and those affecting an entire module.
CC The mutations may be introduced by gene replacement.
XX
XX Sequence 29879 BP; 3672 A; 10596 C; 11545 G; 4066 T; 0 other;
SQ
Query Match 32.3%; Score 32.6; DB 14; Length 29879;
Best Local Similarity 63.3%; Pred. No. 3.9;
Matches 50; Conservative 0; Mismatches 29; Indels 0; Gaps 0;
QY 23 CCCCCAAGGCAAGCTGACGTGATGCGCCGCTCAGCAGCTGGGAATGACTCGTGAT 82
DB 20507 CTCGGCCGACACCTCCCGCAGGCGCAAGCGCTTCAAGACCTCGTTCAGCTCGTGGC 20566
QY 83 GGGGCTAGAGCTGCGCAAC 101
DB 20567 GCGGCTGAGCTGCGCAAC 20585
RESULT 11
AAV21187
ID AAV21187 standard; DNA; 53789 BP.
XX
XX AAV21187;
AC
XX
XX 24-JUL-1998 (first entry)
DT
XX
XX Amycolatopsis mediterranei rifamycin synthesis gene cluster fragment.
DE
XX
XX Amycolatopsis mediterranei rifamycin synthesis; gene cluster;
KW polyketide synthase; actinomycete; ansamycin; ds.
XX
XX Amycolatopsis mediterranei.
OS
XX
XX key Location/Qualifiers
FH CDS 1825..15543
FT /tag= a
FT /label= ORF_A
FT /product= "polyketide synthase"
FT 15550..30759
FT /tag= b

```

```

FT FT /label= ORF_B
FT FT /product= "polyketide synthase"
FT CDS 30895..36060
FT /tag= c
FT /label= ORF_C
FT /product= "polyketide synthase"
FT 36259..41325
FT /tag= d
FT /label= ORF_D
FT /product= "polyketide synthase"
FT 41373..51614
FT /tag= e
FT /label= ORF_E
FT /product= "polyketide synthase"
FT CDS 51713..5293
FT /tag= f
FT /label= ORF_F
FT /product= "polyketide synthase"
XX
XX WO9807868-A1.
XX
XX 26-FEB-1998.
XX
XX 18-AUG-1997; 97MO-EP04495.
XX
XX 20-AUG-1996; 96EP-0810551.
XX
XX (NOVS ) NOVARTIS AG.
XX
XX Engel N, Schupp T, Toupet C;
XX
XX WPI: 1998-169172/15.
DR P-PSDB; AAM52845-W52850.
XX
XX Amycolatopsis mediterranei rifamycin synthesis gene cluster - used
PT to produce rifamycin and rifamycin analogues
XX
XX Claim 4; Page 53-102; 205pp; English.
XX
XX The present sequence represents a Amycolatopsis mediterranei rifamycin
CC synthesis gene cluster DNA fragment from the present invention. The
CC DNA fragment comprises a DNA region involved directly or indirectly
CC in the gene cluster responsible for rifamycin synthesis, including
CC the adjacent DNA regions to the right and left which, by reason of
CC their function in connection with rifamycin biosynthesis, qualify
CC as constituents of this rifamycin gene cluster, and functional
CC fragments, derivatives or constituents of these. The Amycolatopsis
CC mediterranei rifamycin synthesis gene cluster DNA fragment can be used
CC for producing rifamycin, rifamycin analogues or precursors. It can also
CC be used for inactivating or modifying genes involved in ansamycin or
CC rifamycin biosynthesis. The DNA can be used for constructing mutant
CC actinomycetes strains from which the natural rifamycin or ansamycin
CC biosynthesis gene cluster has been partly or completely deleted. The
CC DNA fragment can be used for assembling a library of polyketide
CC synthases, which can be used for assembling a library of polyketides.
CC A hybridisation probe of the invention can be used for identifying DNA
CC fragments involved in the biosynthesis of ansamycins.
XX
XX Sequence 53789 BP; 6707 A; 19183 C; 20504 G; 7395 T; 0 other;
SQ
Query Match 31.9%; Score 32.2; DB 19; Length 53789;
Best Local Similarity 61.2%; Pred. No. 5.3;
Matches 52; Conservative 0; Mismatches 33; Indels 0; Gaps 0;
QY 17 GCGCCTCCCGAAGGCAAGCTGACGTGATGCGCCGCTCAGCAGCTGGGAATGACTC 76
DB 10617 GCACGGGAGCGCGGAGTTTCAGCTGACGCGCTTCAAGACCGGCTTGACCTC 10676
QY 77 GCTGATGGGCTAGAGCTGCGCAAC 101
DB 10677 GCTCAGCGGCTCAGCTGCGCAAC 10701

```

Accession	Gene	Location/Qualifiers
AA14651	AAA14651 standard; DNA; 77536 BP.	
AA14651	AAA14651	
08-AUG-2000	(first entry)	
Nucleotide sequence of the FK-520 biosynthetic gene cluster.		
FK-520	polyketide synthase; PKS; gene cluster; immunosuppressant;	
Streptomyces hygroscopicus var. ascomyceticus; immunophilin;		
FK-506 binding protein; polyketide compound; transplant rejection;		
graft-versus-host disease; uveitis; alopecia universalis;		
autoimmune chronic active hepatitis; inflammatory bowel disease;		
multiple sclerosis; primary biliary cirrhosis; scleroderma;		
neutrite outgrowth; nerve regrowth; Parkinson's disease;		
Alzheimer's disease; stroke; traumatic spinal cord; brain injury;		
peripheral neuropathy; ss.		
Streptomyces hygroscopicus.		
Key	Location/Qualifiers	
CDS	complement (412..1836)	
FT	/*tag- a	
FT	/note- "fkbw gene"	
FT	complement (2020..3579)	
FT	/*tag- b	
FT	/note- "fkbv gene"	
FT	3969..4496	
FT	/*tag- c	
FT	/note- "fkbR2 gene"	
FT	complement (4595..5488)	
FT	/*tag- d	
FT	/note- "fkbR1 gene"	
FT	5601..6818	
FT	/*tag- e	
FT	/note- "fkbE gene"	
FT	6808..8052	
FT	/*tag- f	
FT	/note- "fkbF gene"	
FT	8156..8824	
FT	/*tag- g	
FT	/note- "fkbG gene"	
FT	complement (9122..9883)	
FT	/*tag- h	
FT	/note- "fkbH gene"	
FT	complement (9894..10994)	
FT	/*tag- i	
FT	/note- "fkbI gene"	
FT	complement (10987..11247)	
FT	/*tag- j	
FT	/note- "fkbJ gene"	
FT	complement (11244..12092)	
FT	/*tag- k	
FT	/note- "fkbK gene"	
FT	complement (12113..13150)	
FT	/*tag- l	
FT	/note- "fkbL gene"	
FT	complement (13212..23988)	
FT	/*tag- m	
FT	/note- "fkbC gene"	
FT	complement (13452..13662)	
FT	/*tag- n	
FT	/note- "ACP6"	
FT	complement (13761..14394)	
FT	/*tag- o	
FT	/note- "KR6"	
FT	complement (14517..15294)	
FT	/*tag- p	
FT	/note- "ER6"	
FT	complement (15438..16587)	
FT	/*tag- q	
misc.feature		

FT		/note= "dehydratase domain (DH) 6"
FT	misc_feature	complement (16587..17820)
FT		/*tag- x
FT		/note= "acyltransferase domain (AT) 6"
FT	misc_feature	complement (17820..19053)
FT		/*tag- s
FT	misc_feature	/note= "KS6"
FT		complement (19116..19326)
FT		/*tag- t
FT	misc_feature	/note= "ACP5"
FT		complement (19464..20097)
FT		/*tag- u
FT	misc_feature	/note= "KR5"
FT		complement (20241..21420)
FT		/*tag- v
FT	misc_feature	/note= "DH5"
FT		complement (21420..22653)
FT		/*tag- w
FT	misc_feature	/note= "AT5"
FT		complement (22653..23892)
FT		/*tag- x
FT		/note= "KS5"
FT	CDS	complement (23992..46573)
FT		/*tag- y
FT		/note= "fkbB gene"
FT	misc_feature	complement (24163..24373)
FT		/*tag- z
FT		/note= "ACP4"
FT	misc_feature	complement (24997..26146)
FT		/*tag- aa
FT		/note= "DH4 (inactive)"
FT	misc_feature	complement (26146..27430)
FT		/*tag- ab
FT		/note= "AR4"
FT	misc_feature	complement (27430..28684)
FT		/*tag- ac
FT		/note= "KS4"
FT	misc_feature	complement (28750..28960)
FT		/*tag- ad
FT		/note= "ACP3"
FT	misc_feature	complement (29092..29740)
FT		/*tag- ae
FT		/note= "KR3"
FT	misc_feature	complement (29869..31018)
FT		/*tag- af
FT		/note= "DH3 (inactive)"
FT	misc_feature	complement (31018..32185)
FT		/*tag- ag
FT		/note= "AT3"
FT	misc_feature	complement (32185..33439)
FT		/*tag- ah
FT		/note= "KS3"
FT	misc_feature	complement (33505..33715)
FT		/*tag- ai
FT		/note= "ACP2"
FT	misc_feature	complement (33823..34480)
FT		/*tag- aj
FT		/note= "KR2"
FT	misc_feature	complement (34606..35749)
FT		/*tag- ak
FT	misc_feature	complement (28750..28960)
FT		/*tag- al
FT		/note= "DH2 (inactive)"
FT	misc_feature	complement (35749..37144)
FT		/*tag- am
FT		/note= "AT2"
FT	misc_feature	complement (37145..38296)
FT		/*tag- an
FT		/note= "KS2"
FT	misc_feature	complement (38371..38581)
FT		/*tag- ao
FT		/note= "ACPI"
FT	misc_feature	complement (38677..39307)

```

FT      /tag- ap
FT      /note- "KR1"
FT      complement (39442..40609)
FT      /tag- ag
FT      /note- "DH1"
FT      complement (40609..41842)
FT      /tag- ar
FT      /note- "AT1"
FT      complement (41842..43093)
FT      /tag- as
FT      /note- "KS of extender module 1 (KSI)"
FT      complement (43144..43660)
FT      /tag- at
FT      /note- "ACP of loading domain"
FT      complement (43777..44629)
FT      /tag- au
FT      /note- "ER of loading domain"
FT      complement (44974..46573)
FT      /tag- av
FT      /note- "CoA ligase of loading domain"
FT      46754..47788
FT      /tag- aw
FT      /note- "fkbp gene"
FT      47785..52272
FT      /tag- ax
FT      /note- "fkbp gene"
FT      52275..71465
FT      /tag- ay
FT      /note- "fkba gene"
FT      52362..53576
FT      /tag- az
FT      /note- "KS7"
FT      53577..54716
FT      /tag- bb
FT      /note- "AF7"
FT      54717..55871
FT      /tag- ba
FT      /note- "DH7"
FT      56019..56819
FT      /tag- bc
FT      /note- "ER7"
FT      56943..57575
FT      /tag- bd
FT      /note- "KR7"
FT      57710..57920
FT      /tag- be
FT      /note- "ACP7"
FT      57990..59243
FT      /tag- bf
FT      /note- "KS8"
FT      59244..60398
FT      /tag- bg
FT      /note- "AT8"
FT      60399..61412
FT      /tag- bh
FT      /note- "DH (inactive)"

Query Match      31.3%: Score 31.6; DB 21; Length 77536;
Best Local Similarity 65.7%; Pred. No. 8.1; Mismatches 24; Indels 0; Gaps 0;
Matches 46; Conservative 0;

QY      32 CAAGCTGACGATGATCGCGCTCACGAGCCCTGGAATGAGCTCGCTGATGGGCTAGA 91
      |||||
DB      62384 CGAGATCGCGCGCACGACGCTTCAAGACCTCGGATGAGACTCGCTCACCGCGATCGA 62443
      |||||
QY      92 GCTGCGCAC 101
      |||||
DB      62444 GCTGCGCAC 62453

```

RESULT 13
 AAA14663
 ID AAA14663 standard; DNA; 4466 BP.

```

XX      AAA14663;
AC      08-AUG-2000 (first entry)
XX      Nucleotide sequence of a fragment of the FK-520 PKS gene cluster.
DT      FK-520; polyketide synthase; PKS; gene cluster; immunosuppressant;
XX      immunophilin; FK-506 binding protein; polyketide compound; uveitis;
DT      transplant rejection; graft-versus-host disease; alopecia universalis;
XX      autoimmune chronic active hepatitis; inflammatory bowel disease;
DE      multiple sclerosis; primary biliary cirrhosis; scleroderma;
XX      neurite outgrowth; nerve regrowth; Parkinson's disease;
KW      Alzheimer's disease; stroke; traumatic spinal cord; brain injury;
KM      peripheral neuropathy; ss.
XX      Synthetic.
OS      Streptomyces hygroscopicus.
XX      Key      Location/Qualifiers
FH      CDS      9..4454
FT      /tag- a
FT      /note- "no termination codon given"
PN      WO200020601-A2.
PD      13-APR-2000.
XX      01-OCT-1999; 99WO-US22886.
XX      02-OCT-1998; 98OS-0102748.
XX      11-MAR-1999; 99US-0123810.
XX      17-JUN-1999; 99US-0139650.
PA      (KOSA-) KOSAN BIOSCIENCES INC.
XX      Reeves C, Chu D, Khosla C, Santl D, Wu K;
PI      MPI: 2000-317716/27.
DR      P-PDB; AAY84727.
XX      New isolated polyketide synthase nucleic acid and polyketide compounds,
PT      useful for treating e.g. transplant rejection, uveitis, multiple
PT      sclerosis, Alzheimer's disease, Parkinson's disease, stroke, or
PT      peripheral neuropathy -
XX      Example 1; Page 82-85; 126pp; English.
PS      The present sequence represents an NheII-XhoII fragment that encodes
XX      module 8 of the FK-520 polyketide synthase (PKS) gene cluster with the
XX      endogenous acyltransferase (AT) domain replaced by the AT domain of
CC      module 12 of the rapamycin PKS (which is specific for malonyl CoA).
CC      FK-506 is a potent immunosuppressant, and acts through initial formation
CC      of an intermediate complex with protein immunophilins known as FK-506
CC      binding proteins. The nucleic acids are used for producing polyketide
CC      compounds. The polyketide compounds can be used as immunosuppressants to
CC      prevent or treat transplant rejection, graft-versus-host disease or
CC      uveitis. They can also be used for treating e.g. alopecia universalis,
CC      autoimmune chronic active hepatitis, inflammatory bowel disease,
CC      multiple sclerosis, primary biliary cirrhosis, or scleroderma. They
CC      also have neurotrophic activity and can be used to promote neurite
CC      outgrowth in NGF-treated PC12 cells and in sensory neuronal cultures,
CC      and in intact animals, they promote regrowth of damaged facial and
CC      sciatic nerves, and repair lesioned serotonin and dopamine neurons in
CC      the brain. They can also be used for treating e.g. Parkinson's disease,
CC      Alzheimer's disease, stroke, traumatic spinal cord and brain injury, or
CC      peripheral neuropathies. They can also be used in agricultural and
CC      veterinary applications.
XX      Sequence 4466 BP; 634 A; 1796 C; 1433 G; 603 T; 0 other;
SQ      Query Match      31.1%: Score 31.4; DB 21; Length 4466;
      Best Local Similarity 67.7%; Pred. No. 7.2;

```


XX Example 1: Page 85-88; 126pp; English.

CC The present sequence represents an Nherl-xhoII fragment that encodes
 CC module 8 of te FK-520 polyketide synthase (PKS) gene cluster with the
 CC endogenous acyltransferase (AT) domain replaced by the AT domain of
 CC module 13 of the rapamycin PKS (which is specific for methylmalonyl CoA).
 CC FK-506 is a potent immunosuppressant, and acts through initial formation
 CC of an intermediate complex with protein immunophilins known as FK-506
 CC binding proteins. The nucleic acids are used for producing polyketide
 CC compounds. The polyketide compounds can be used as immunosuppressants to
 CC prevent or treat transplant rejection, graft-versus-host disease or
 CC uveitis. They can also be used for treating e.g. alopecia universalis,
 CC autoimmune chronic active hepatitis, inflammatory bowel disease,
 CC multiple sclerosis, primary biliary cirrhosis, or scleroderma. They
 CC also have neurotrophic activity and can be used to promote neurite
 CC outgrowth in NGF-treated PC12 cells and in sensory neuronal cultures,
 CC and in intact animals, they promote regrowth of damaged facial and
 CC sciatic nerves, and repair lesioned serotonln and dopamine neurons in
 CC the brain. They can also be used for treating e.g. Parkinson's disease,
 CC Alzheimer's disease, stroke, traumatic spinal cord and brain injury, or
 CC peripheral neuropathies. They can also be used in agricultural and
 CC veterinary applications.

XX Sequence 4547 BP; 658 A; 1851 C; 1437 G; 601 T; 0 other;

Query Match 31.1%; Score 31.4; DB 21; Length 4547;

Best Local Similarity 67.7%; Pred. No. 7.2; Mismatches 0; Gaps 0;

DB 37 TCGAGCTGATGCGCGCCGCTCAGAGCCTGGGAATGACATCGCTGATGGGCTAGAGCTGC 96
 85 TCCCGCGAGCGCGGCGGTTCAAGAGACCTCGCATCGACTCGCTACCGCGGTCACGCTGC 144

QY 97 GCAAC 101

DB 145 GCAAC 149

Search completed: November 6, 2002, 13:16:46
 Job time : 255 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 6, 2002, 11:51:49 : Search time 443.625 Seconds

(without alignments)
6625.829 Million cell updates/sec

Title: US-09-724-876-2_COPY_37570_37670

Perfect score: 101

Sequence: 1 aggtctgcagcgtctgcgc.....tgggctagagctgcgcgaac 101

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :
1: gb_da:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: gb_da:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_ph:*
24: em_pat:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rtd:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	101	100.0	58733	1 AF217189	AF217189 Sorangium
2	101	100.0	71989	6 AR172664	AR172664 Sequence
3	99.4	98.4	68750	1 AF210843	AF210843 Sorangium
4	99.4	98.4	68750	6 AR193029	AR193029 Sequence
5	99.4	98.4	68750	6 AR199551	AR199551 Sequence
6	99.4	98.4	68750	6 AR199559	AR199559 Sequence
7	99.4	98.4	68750	6 AR199557	AR199557 Sequence
8	99.4	98.4	68750	6 AR201097	AR201097 Sequence
9	99.4	98.4	68750	6 AR208671	AR208671 Sequence
10	80.2	79.4	16124	1 AX024384	AX024384 Sequence
11	80.2	79.4	16124	6 AX024277	AX024277 Sequence
12	49.2	48.7	978	1 AX024375	AX024375 Sequence
13	49.2	48.7	978	6 AX024268	AX024268 Sequence
14	45.4	45.0	42603	1 AF188287	AF188287 Stigmatel
15	43.6	43.2	82746	1 AF453501	AF453501 Actinosyn
16	40.4	40.0	66808	1 SAU421825	SAU421825 Stigmatel
17	40.2	39.8	5292	6 AX153793	AX153793 Sequence
18	40.2	39.8	34071	6 AX153790	AX153790 Sequence
19	40.2	39.8	42717	6 AX153789	AX153789 Sequence
20	40.2	39.8	49736	1 AF319998	AF319998 Stigmatel
21	38.6	38.2	30000	6 AX250261	AX250261 Sequence
22	37.6	37.2	105495	2 AC114475	AC114475 Magnaport
23	35.4	35.0	104326	1 AB070940	AB070940 Streptomy
24	35.2	34.9	33529	6 AR166425	AR166425 Sequence
25	35.2	34.9	105495	2 AC114475	AC114475 Magnaport
26	34.4	34.1	14874	1 AE005927	AE005927 Caulobact
27	34.2	33.9	4615	6 AX153791	AX153791 Sequence
28	33.8	33.5	11890	1 MTCY275	295617 Mycobacteri
29	33.8	33.5	14980	1 SC4A9	AL13668 Streptomy
30	33.8	33.5	18614	1 AE007033	AE007033 Mycobacte
31	33.8	33.5	30000	6 AX250262	AX250262 Sequence
32	33.6	33.3	14402	1 AF098795	AF098795 Pseudomon
33	33.3	32.7	1721	8 AF488522	AF488522 Oryza sat
34	33.3	32.7	1784	1 AF411573S3	AF411575 Actinomad
35	33.3	32.7	1791	1 PST249741	AJ249741 Pseudomon
36	33.3	32.7	6564	6 AX069073	AX069073 Sequence
37	33.3	32.7	20679	1 MTCY19H9	283857 Mycobacteri
38	33.3	32.7	32870	1 AE007122	AE007122 Mycobacte
39	33.3	32.7	107379	1 SHGCP1R	X86780 S. hygroscop
40	33.3	32.7	112133	2 AP005254	AP005254 Oryza sat
41	32.8	32.5	43280	1 SFU78289	U78289 Streptomyce
42	32.6	32.3	20235	1 SPREXYAB	M63677 S. erythraea
43	32.6	32.3	20235	6 AR049368	AR049368 Sequence
44	32.6	32.3	20235	6 AR095529	AR095529 Sequence
45	32.6	32.3	20444	1 SBERRYABS	X62569 S. erythraea

ALIGNMENTS

RESULT 1
LOCUS AF217189
DEFINITION AF217189 58733 bp DNA linear BCT 09-JUN-2000
Sorangium cellulosum putative transposase gene, partial cds;
complete cds.
ACCESSION AF217189
VERSION AF217189
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1 (bases 1 to 58733)
Sorangium cellulosum.
Polyangium cellulosum
Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
Myxococcales; Sorangineae; Polyangiaceae; Polyangium.

Pred. No. is the number of results predicted by chance to have a

QSTRLEWETTLIDLKRLPMLGDHROGAVNPPGAAYLEMAISSGAELDGPLOITDVA
LAELAFAAGDAVAVOYVTTTEPQSPGRLOFOJASRAGRGHASFVHARGALLRVERTE
VPAQITLISAVARLQASMPAAATTAETEMKLOJQGPAGOGIAELMRGGLALGVRLPE
DAAGSAEIRLHPALDNCFOYVGSJLFRAGGEATPWVEVGSRLQRPGEIMCHRA
RVNHHGQTPDRGADFWYDSSGAAVVAEVLVAORLPJGVRRREEDBNLEWEP
AAVGTAKVNRMLLGGGGGJGALRSMLEAGHAAVVAHAESTSSAAGVALLAKAF
DGOPTAVHVLGSLDGGGELDPGAGALDAPARSADVSPALDPAIVRGDDSLMTV
OALGMGFPRADPRMLTLTRGQAVGAGDVSVTOAPLIGISRVITMERADFCARVDL
PTPRDGLGALIAELLDADAEPVALRGECVARYTRROPTPRRISCVPTDVT
IRADSTLVTVGGJLGLSVAGMLAERGAGHLVYGRSGAASVQRAVAALEARGAR
VTAKADAVNQLERILREVTTISGMLRGVNAAGLIDDLMOQPPARPRKVMAP
VOGALHHLHALTREALSEFVLYASVGLSPGGGNAATATFEDALAHHRAGLPVA
LSYDMGLFAEYGMMAAOEDRGARLVSRGMRSLTPTDEGLSALRLESVRGDDVPM
PLMVELYPPAASRMLSLRTVTAHRAASGAGPDGDLRLRAAEPARSGLLEPRLR
AOISQVIRLEPGRKEVDAPLITSLGMSMLGELNRRLEAMGITTVPATLMTYPTVAA
LSGHLAREACBAEPVESPTTADSAVELEEMSDTLTOLIAFKKALT"

gene
16134. .37907
/gene="epod"
16134. .37907
/gene="epod"
/note="type 1 polyketide synthase; epothilone PKS modules
3, 4, 5, and 6"
/codon_start=1
/transl_table=1
/product="epod"
/protein_id="AAF62883.1"
/db_xref="GI:7453560"
/translation="MTTRGPTAQQNPILKQAIIITQRLERLAIQAELERTPIATV
GIGCFRPGADAPAEFWELLDAERDAVOPLDRMALVGAVEAVPMWAGLTPIDC
FDAFPGISPREARSLDPOHRLLEVAWEGLEDJIPRSIDGSRCTVFGALFYADA
RYVARLPRERDAVSATGNMLSIAGRLSTJLGOGPCLVYDTCSSSLVAILHAGS
LRAGESDLALAGVSTLSPDMMEAAARTALSPDRCGRPTDASNFANFREGGILVY
LKRISDAQRGDRIMALIRGSATIHNDGRSTGTRPNTLAQETVYREALRSRNVAGAY
DYVTEHGTGSLDPIEVEALRATVPARSQTCVIGAVTNTIGHLEAAGVGLIK
AAULTEHRIIPRNLFNPRIRLESGSALALATEPVMPRTDPRRAGVSGFSMGST
NAHVLEEAPEAVELMPAPAPERSAELVLSGSEBALDQAARLEHDMHELGLGV
AFSLATRSAMTHRLAAVATSRREGILALVSGAGGTPAGARCIASSRKLALFT
GOGAQTGPMGRGLCAAMPAPAREARDRCVTFLEDELDPRLCPVMAFAGASLLDPT
AFQORAFAYEYALTAIAMSNGVPELLVHSIGELYACVAFSLEDGRVLYAAGS
RLMOGLSAGAMVSLGAPLEVAVAAPVPAHWVSTIAVNGEVOYVINGEDAVALIA
GFAARGVTRKRLVSHAFSHPLMEPLMEFGRAVASVYTRRPSVSLVNSLKGVTDE

Query Match 100.0%; Score 101; DB 1; Length 58733;
Best Local Similarity 100.0%; Pred. No. 5,5e-15;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGCTCGCAGGTGCTGCCCTCCCGAAGCAAGCTGACGTGATGCCGCGTACGA 60
|||||
Db 37570 AGGCTCGCAGGTGCTGCCCTCCCGAAGCAAGCTGACGTGATGCCGCGTACGA 37629
|||||
QY 61 GCCTGGGAATGACCTGCTGATGGGGCTAGAGCTGCCGAAC 101
|||||
Db 37630 GCCTGGGAATGACCTGCTGATGGGGCTAGAGCTGCCGAAC 37670
|||||

RESULT 2
LOCUS AR172664 71989 bp DNA linear PAT 17-DEC-2001
DEFINITION Sequence 2 from patent US 6303342.
ACCESSION AR172664
VERSION AR172664.1 GI:17912155
KEYWORDS
SOURCE .
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 71989)
AUTHORS Julien,B., Katz,L., Khosla,C. and Tang,L.
TITLE Recombinant methods and materials for producing epothilones C and D
JOURNAL Patent: US 6303342-A 2 16-OCT-2001;
FEATURES
Location/Qualifiers
1. .71989
/organism="unknown"
BASE COUNT 10108 a 23531 c 26617 g 11731 t 2 others

ORIGIN

Query Match 100.0%; Score 101; DB 6; Length 71989;
Best Local Similarity 100.0%; Pred. No. 5,4e-15;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGCTCGCAGGTGCTGCCCTCCCGAAGCAAGCTGACGTGATGCCGCGTACGA 60
|||||
Db 37570 AGGCTCGCAGGTGCTGCCCTCCCGAAGCAAGCTGACGTGATGCCGCGTACGA 37629
|||||
QY 61 GCCTGGGAATGACCTGCTGATGGGGCTAGAGCTGCCGAAC 101
|||||
Db 37630 GCCTGGGAATGACCTGCTGATGGGGCTAGAGCTGCCGAAC 37670
|||||

RESULT 3
LOCUS AF210843 68750 bp DNA linear BCT 21-JAN-2000
DEFINITION Sorangium cellulosum strain So ce90 epothilone biosynthesis gene
cluster, complete sequence.
ACCESSION AF210843
VERSION AF210843.1 GI:6724237
KEYWORDS
SOURCE .
ORGANISM Polyangium cellulosum.
Polyangium cellulosum.
Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
Mycococcales; Sorangineae; Polyangiaceae; Polyangium.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Molnar,I., Schupp,T., Ono,M., Zirkle,R., Milnamow,M.,
Nowak-Thompson,B., Engel,N., Toupet,C., Strattmann,A., Cyr,D.D.,
Gorlach,J., Mayo,J.M., Hu,A., Goff,S., Schmid,J. and Ligon,J.M.
TITLE The biosynthetic gene cluster for the microtubule-stabilizing
agents epothilones A and B from Sorangium cellulosum So ce90
Chem. Biol. 7 (2), 97-109 (2000)

JOURNAL
MEDLINE PubMed
20130945
10662695
REFERENCE 2 (bases 1 to 68750)
AUTHORS Molnar,I.
TITLE Direct Submission
JOURNAL Submitted (03-DEC-1999) Natural Product Genetics, Novartis
Business Research Institute, Inc., 3054 Cornwallis Rd, P.O.Box
12257, Research Triangle Park, NC 27709, USA

FEATURES
source
1. 68750
Location/Qualifiers
/organism="Polyangium cellulosum"
/strain="So ce90"
/db_xref="taxon:56"
<1. .1826
/note="ORF1"
/codon_start=3
/transl_table=1
/product="unknown"
/protein_id="AAF26904.1"
/db_xref="GI:6724238"
/translation="ASLDALFARATSAVRLVDHGRGATREHVLAENGRIEDLARREH
LRIDGSPRECHKLGIDLYVELLHNDOPLASISHHARSLRHPRWSTDAMLYGPRAY
RWLAKAPGCPLEIREBERERAKTAQERLRMLAAAPCEAPDLPREDNANGLPB
MSPEVAERLRISYATPELACALALAWGTGPGWGPAYEMLEENLLGFLPT
AIAAASAPGTSEALRGAARLFAEMEVVSSKSGQNGNIPEALMERLRTIYRAMGNAD
LSFERAEALIAEVRRLAOPAPPAAGAGLAVAGSSGRLSGLVTDGDAVSGDND
IVMFOGRISVVVLVAGNDPEELAPLSQMLFAVHNAAGTISKVLTEGSPLIYMAN
QARPMISVHARGFAMVNOAMVPPREGAPRVYORSTIMEHEHPTPCHEPAGSAS
LACDEHLVCELSAGRLERKRHHKRGASRAYLGEHIAATWTPSFLNATHVL
WADPDRAALIGVDKRTVEPIVLAETHRPAHVASEDRDIFALGDPDSRWVHEHRL
SGASTVADYORQLMDRDPDWVLRGLFTTNDRIILYARS"

CDS
1 (bases 1 to 68750)
complete(1900. .3171)
/note="ORF2: similar to hypothetical proteins from
Mycobacterium tuberculosis and streptomyces coelicolor and
to DD-peptidases"
/codon_start=1
/transl_table=1
/product="unknown"
/protein_id="AAF26905.1"

Best Local Similarity 99.0%; Pred. No. 1.3e-14;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AGGTCGACAGTGTGGGCTCCCGAAGGCAAGCTGAGTGGCGGCTCAGCA 60
|||||
Db 43182 AGGTCGACAGTGTGGGCTCCCGAAGGCAAGCTGAGTGGCGGCTCAGCA 43241

Qy 61 GCCTGGGAATGAGCTCGCTGATGGGGCTAGAGCTCGGCAAC 101
|||||
Db 43242 GCCTGGGAATGAGCTCGCTGATGGGGCTAGAGCTCGGCAAC 43282

RESULT 4
LOCUS AR193029 68750 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 1 from patent US 6346404.
ACCESSION AR193029
VERSION AR193029.1 GI:20238994
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ilyon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and Gorlach,J.
TITLE Genes for the biosynthesis of epoethilones
JOURNAL Patent: US 6346404-A 1 12-FEB-2002;
FEATURES
source 1..68750
/organism="unknown"
ORIGIN

BASE COUNT 9596 a 22456 c 25539 g 11159 t

Query Match 98.4%; Score 99.4; DB 6; Length 68750;
Best Local Similarity 99.0%; Pred. No. 1.3e-14;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AGGTCGACAGTGTGGGCTCCCGAAGGCAAGCTGAGTGGCGGCTCAGCA 60
|||||
Db 43182 AGGTCGACAGTGTGGGCTCCCGAAGGCAAGCTGAGTGGCGGCTCAGCA 43241

Qy 61 GCCTGGGAATGAGCTCGCTGATGGGGCTAGAGCTCGGCAAC 101
|||||
Db 43242 GCCTGGGAATGAGCTCGCTGATGGGGCTAGAGCTCGGCAAC 43282

RESULT 5
LOCUS AR199551 68750 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 1 from patent US 6355457.
ACCESSION AR199551
VERSION AR199551.1 GI:20249625
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ilyon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and Gorlach,J.
TITLE Genes for the biosynthesis of epoethilones
JOURNAL Patent: US 6355457-A 1 12-MAR-2002;
FEATURES
source 1..68750
/organism="unknown"
ORIGIN

BASE COUNT 9596 a 22456 c 25539 g 11159 t

Query Match 98.4%; Score 99.4; DB 6; Length 68750;
Best Local Similarity 99.0%; Pred. No. 1.3e-14;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AGGTCGACAGTGTGGGCTCCCGAAGGCAAGCTGAGTGGCGGCTCAGCA 60
|||||
Db 43182 AGGTCGACAGTGTGGGCTCCCGAAGGCAAGCTGAGTGGCGGCTCAGCA 43241

Db 43182 AGGTCGACAGTGTGGGCTCCCGAAGGCAAGCTGAGTGGCGGCTCAGCA 43241

Qy 61 GCCTGGGAATGAGCTCGCTGATGGGGCTAGAGCTCGGCAAC 101
|||||
Db 43242 GCCTGGGAATGAGCTCGCTGATGGGGCTAGAGCTCGGCAAC 43282

RESULT 6
LOCUS AR199559 68750 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 1 from patent US 6355458.
ACCESSION AR199559
VERSION AR199559.1 GI:20249633
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ilyon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and Gorlach,J.
TITLE Genes for the biosynthesis of epoethilones
JOURNAL Patent: US 6355458-A 1 12-MAR-2002;
FEATURES
source 1..68750
/organism="unknown"
ORIGIN

BASE COUNT 9596 a 22456 c 25539 g 11159 t

Query Match 98.4%; Score 99.4; DB 6; Length 68750;
Best Local Similarity 99.0%; Pred. No. 1.3e-14;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AGGTCGACAGTGTGGGCTCCCGAAGGCAAGCTGAGTGGCGGCTCAGCA 60
|||||
Db 43182 AGGTCGACAGTGTGGGCTCCCGAAGGCAAGCTGAGTGGCGGCTCAGCA 43241

Qy 61 GCCTGGGAATGAGCTCGCTGATGGGGCTAGAGCTCGGCAAC 101
|||||
Db 43242 GCCTGGGAATGAGCTCGCTGATGGGGCTAGAGCTCGGCAAC 43282

RESULT 7
LOCUS AR199567 68750 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 1 from patent US 6355459.
ACCESSION AR199567
VERSION AR199567.1 GI:20249641
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ilyon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and Gorlach,J.
TITLE Genes for the biosynthesis of epoethilones
JOURNAL Patent: US 6355459-A 1 12-MAR-2002;
FEATURES
source 1..68750
/organism="unknown"
ORIGIN

BASE COUNT 9596 a 22456 c 25539 g 11159 t

Query Match 98.4%; Score 99.4; DB 6; Length 68750;
Best Local Similarity 99.0%; Pred. No. 1.3e-14;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AGGTCGACAGTGTGGGCTCCCGAAGGCAAGCTGAGTGGCGGCTCAGCA 60
|||||
Db 43182 AGGTCGACAGTGTGGGCTCCCGAAGGCAAGCTGAGTGGCGGCTCAGCA 43241

Qy 61 GCCTGGGAATGAGCTCGCTGATGGGGCTAGAGCTCGGCAAC 101
|||||
Db 43242 GCCTGGGAATGAGCTCGCTGATGGGGCTAGAGCTCGGCAAC 43282

RESULT 8
LOCUS AR201097 68750 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 1 from patent US 6358719.
ACCESSION AR201097
VERSION AR201097.1 GI:20251985
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and Goriach,J.
TITLE Genes for the biosynthesis of epothilones
JOURNAL Patent: US 6358719-A 1 19-MAR-2002;
FEATURES
Location/Qualifiers
source 1..68750
/organism="unknown"
BASE COUNT 9596 a 22456 c 25539 g 11159 t
ORIGIN
Query Match 98.4%; Score 99.4; DB 6; Length 68750;
Best Local Similarity 99.0%; Pred. No. 1.3e-14;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 AGGTCGCGAGGTGTCGCCCTCCCGAAGCGCAAGCTCGACGTGATGCGCCGCTACGA 60
|||||
Db 43182 AGGTCGCGAGGTGTCGCCCTCCCGAAGCGCAAGCTCGACGTGATGCGCCGCTACGA 43241
|||||
OY 61 GCCTGGGAATGACACTGCTGATGGGGCTAGAGCTGCGCAAC 101
|||||
Db 43242 GCCTGGGAATGACACTGCTGATGGGGCTAGAGCTGCGCAAC 43282
|||||
RESULT 9
LOCUS AR208671 68750 bp DNA linear PAT 20-JUN-2002
DEFINITION Sequence 1 from patent US 6383787.
ACCESSION AR208671
VERSION AR208671.1 GI:21509886
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and Goriach,J.
TITLE Genes for the biosynthesis of epothilones
JOURNAL Patent: US 6383787-A 1 07-MAY-2002;
FEATURES
Location/Qualifiers
source 1..68750
/organism="unknown"
BASE COUNT 9596 a 22456 c 25539 g 11159 t
ORIGIN
Query Match 98.4%; Score 99.4; DB 6; Length 68750;
Best Local Similarity 99.0%; Pred. No. 1.3e-14;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 AGGTCGCGAGGTGTCGCCCTCCCGAAGCGCAAGCTCGACGTGATGCGCCGCTACGA 60
|||||
Db 43182 AGGTCGCGAGGTGTCGCCCTCCCGAAGCGCAAGCTCGACGTGATGCGCCGCTACGA 43241
|||||
OY 61 GCCTGGGAATGACACTGCTGATGGGGCTAGAGCTGCGCAAC 101
|||||
Db 43242 GCCTGGGAATGACACTGCTGATGGGGCTAGAGCTGCGCAAC 43282
|||||
RESULT 10
LOCUS AX024384 16124 bp DNA linear BCT 15-SEP-2000
AX024384

DEFINITION Sequence 82 from Patent DE19846493.
ACCESSION AX024384
VERSION AX024384.1 GI:10184588
KEYWORDS
SOURCE Polyangium cellulosum.
ORGANISM Polyangium cellulosum
Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
Myxococcales; Sorangineae; Polyangiaceae; Polyangium.
REFERENCE 1 (bases 1 to 16124)
AUTHORS Beyer,S. and Mueller,R.J.
JOURNAL Patent: DE 19846493-A 13-APR-2000;
FEATURES
Location/Qualifiers
source 1..16124
/organism="Polyangium cellulosum"
/db_xref="taxon:56"
BASE COUNT 2336 a 5400 c 5695 g 2693 t
ORIGIN
Query Match 79.4%; Score 80.2; DB 1; Length 16124;
Best Local Similarity 87.1%; Pred. No. 6.9e-10;
Matches 88; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
OY 1 AGGTCGCGAGGTGTCGCCCTCCCGAAGCGCAAGCTCGACGTGATGCGCCGCTACGA 60
|||||
Db 5947 AGGTCGCGAGGTGTCGCCCTCCCGAAGCGCAAGCTCGACGTGATGCGCCGCTACGA 6006
|||||
OY 61 GCCTGGGAATGACACTGCTGATGGGGCTAGAGCTGCGCAAC 101
|||||
Db 6007 GCATGGCATGACACTGCTGATGAGCCTGAGAGCTGCGCAAC 6047
|||||

RESULT 11
LOCUS AX024277 16124 bp DNA linear PAT 15-SEP-2000
DEFINITION Sequence 82 from Patent DE19846493.
ACCESSION AX024277
VERSION AX024277.1 GI:10184551
KEYWORDS
SOURCE Polyangium cellulosum.
ORGANISM Polyangium cellulosum
Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
Myxococcales; Sorangineae; Polyangiaceae; Polyangium.
REFERENCE 1 (bases 1 to 16124)
AUTHORS Beyer,S. and Mueller,R.J.
JOURNAL Patent: DE 19846493-A 82 13-APR-2000;
FEATURES
Location/Qualifiers
source 1..16124
/organism="Polyangium cellulosum"
/db_xref="taxon:56"
BASE COUNT 2336 a 5400 c 5695 g 2693 t
ORIGIN
Query Match 79.4%; Score 80.2; DB 6; Length 16124;
Best Local Similarity 87.1%; Pred. No. 6.9e-10;
Matches 88; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
OY 1 AGGTCGCGAGGTGTCGCCCTCCCGAAGCGCAAGCTCGACGTGATGCGCCGCTACGA 60
|||||
Db 5947 AGGTCGCGAGGTGTCGCCCTCCCGAAGCGCAAGCTCGACGTGATGCGCCGCTACGA 6006
|||||
OY 61 GCCTGGGAATGACACTGCTGATGGGGCTAGAGCTGCGCAAC 101
|||||
Db 6007 GCATGGCATGACACTGCTGATGAGCCTGAGAGCTGCGCAAC 6047
|||||
RESULT 12
LOCUS AX024375 978 bp DNA linear BCT 15-SEP-2000
DEFINITION Sequence 73 from Patent DE19846493.
ACCESSION AX024375
VERSION AX024375.1 GI:10184579

KEYWORDS	
SOURCE	Polyangium cellulorum.
ORGANISM	Polyangium cellulorum.
REFERENCE	Bacteria; Proteobacteria; delta subdivision; Myxobacteria; Mycococciales; Sorangiineae; Polyangiaceae; Polyangium. 1 (bases 1 to 978)
AUTHORS	Beyer,S. and Mueller,R.J.
JOURNAL	Patent: DE 19846493-A, 13-APR-2000;
FEATURES	BIOTECHNOLOG FORSCHUNG GMBH (DE) Location/Qualifiers 1..978
source	/organism="Polyangium cellulorum" /db_xref="taxon:56"
BASE COUNT	154 a 346 c 307 g 171 t
ORIGIN	
Query Match	48.7%; Score 49.2; DB I; Length 978;
Best Local Similarity	70.2%; Pred. No.0.03;
Matches	66; Conservative 0; Mismatches 28; Indels 0; Gaps 0;
OY	8 GCAGTGCTGCGCCCTCCCCGAAGCACAACGTGCATGGCGCGTCACGAGCCTGSG 67 Db 808 GCGGGGCGCTGCCCTTCCGCCGCGAAGATGCAAGCGCACGTCGCGTACAGAATCTCGG 749 OY 68 AATGACACTCGATAGGGGCTAGAGCTCGCAAC 101 Db 748 CATGACACTCGCATAGGCGCTGAGCTCCGCAAC 715

[illegible]

REFERENCE	AUTHORS	TITLE
1 (bases 1 to 42603)	Slakowski,B., Schaiter,H.U., Ehret,H., Kunze,B., Weinig,S., Nordiek,G., Brandt,P., Bloeker,H., Hofle,G., Beyer,S. and Muller,R.	New lessons for combinatorial biosynthesis from myxobacteria. The myxothiazol biosynthetic gene cluster of <i>Stigmatella aurantiaca</i> DM4/3-1
JOURNAL MEDLINE PUBMED	J. Biol. Chem. 274 (52), 37391-37399 (1999)	
2 (bases 1 to 42603)	Slakowski,B., Schaiter,H.U., Ehret,H., Kunze,B., Weinig,S., Nordiek,G., Brandt,P., Bloeker,H., Hofle,G., Beyer,S. and Mueller,R.	
JOURNAL	Submitted (20-SEP-1999) NBI/MX, GBF, Mascheroder Weg 1, Braunschweig 38124, Germany	
FEATURES	Location/Qualifiers	
source	1. 42603	
	/organism="Stigmatella aurantiaca"	
	/strain="DM4/3-1"	
	/db_xref="taxon:41"	
	1. 834	
	/gene="mtaA"	
gene	1. 834	
	/gene="mtaA"	
	/note="pant transferase"	
	/codon_start=1	
	/transl_table=11	
	/product="MtaA"	
	/protein_id="AAI19809.1"	
	/db_xref="GI:6635394"	
	/translation="MPTSSPALPLKLPPDEYVWVIIVEPINEPINEGRIETALLDPG ERKQORFYERERHRIQYVNSHALVRLTISAYAPAEANSFSANOGRPIREBKPW LRNLSHTDGMALCAVARDDVGADVEDERKEVEIADSFPAEVAASLRALPYSG QRERFDFYWLTKEAYI1KARGMGLSLPDDFAFVSGLSRTISFDPLVDPSQWQFV RFPFSORHAAALVARRRSEAPLTVRFQRTVPLODDAPAEYLSAERLOPRLRMQVGG G"	
	1388. 13399	
	/gene="mtab"	
	/note="PKS"	
	1388. 13399	
	/gene="mtab"	
	/note="PKS: Mtab loading domain and module 1: polyketide synthase domains: ACP, KS, AT, AT, DH, ER, KR, ACP, Mtab module 2: polyketide synthase domains: KS, AT, DH, KR, ACP"	
	/codon_start=1	
	/transl_table=11	
	/product="Mtab"	
	/protein_id="AAI19810.1"	
	/db_xref="GI:6635395"	
	/translation="MSTEHGQVGVGRSTQGLRDMLYQOVASOYGLBAHTVRITEPFS LIGSVYFATRLIAELARWLERPLSPILVWENHPTIDKFARHLSGAEPRLEMAHGCHEA AVGADPIPAIVGLACREPKAPDAKAFWRYLQEGDCVTEVPSRRMDASLYDPAPGP GKIVTVEYGAFLDSVDRPDPEFGISPRREGAGLDLPQORIMELISMEALPDAGIAPRLR GSRGTGFTGMWRDYNADLOTIHNAAPTLHRTCTYAGGIITANRISVALGEGSLAIDY ACSSLSVAIHEACRSLSRGESSLALAGVQVLNTPPTFMAVSAFGLADPGRCRTIDA RASGVYRGGCGVLYLKLPSLRALADDAIYCLIRGSVAVNDGASNLCTANPAPQGR LRPAVYVAGMDPAQVHVVELHGTGTPLGPITENGALGAVYLGAERERPLHVSQVSTN IGLLEAAAGIASIVYKVALSIRHOQLPASIHFEQNPRIPTDRLRLVQGLAPWMPKE DPVAVATISAFEGCTGCHVYVEHRRGSGSTRAAPVSTPAHOSRHHKTAFLFGACASA WMGMGELLIGAEYFYPARRECECDAAFRPLGAGNSLYVALLAPRPSRFEQGRPLVLA AFOISIALALMRSGIPEPSATVIGYSVGEAAAOVAGILITIEVFALAVHYI1QVQKGV GYGVMAVYVAGSIAEVAERIAPIYGGRIISASNGSRVGLGGEAALODFTASMTAAGIF ARVLEIDFAYIALPOVALKPLFLOSCPVLRSPGRVPSVSGAFVDGGTIGAEYWA NSNREVPFATAMRTLLEGGYDVLDVNPPIIHEHSVLEWVRAGTASAEFISMRGE PSVEGIEHRAILOTLSOAGSPYRWRERVPYGTGAGSSAPVTEIIPVARSARTOALREOADR IAOQLRDAASSLTIDCTYAGARSHHEHRLIAYATRALALEGLAAPFARGETRPPS QGNVARGEAPRIGVFESGGGSPFGSSVOLFEEFESFPAVVEGSEKRLFRQGTASLIE ALHAEESESHPPTEIEOPLFLFYQVALALHMSWQIRDAVWSSERLIPVISAHVH	

REFERENCE 1 Actinosynnema.
AUTHORS Yu,T.W., Bai,L., Clade,D., Hoffmann,D., Toelzer,S., Trinh,K.Q.,
TITLE Xu,J., Moss,S.J., Leistner,E. and Floss,H.G.
JOURNALS The biosynthetic gene cluster of the maytansinoid antitumor agent
MEDLINE ansamitocin from Actinosynnemapretiosum
22056096 Proc. Natl. Acad. Sci. U.S.A. 99 (12), 7968-7973 (2002)
PUBMED 12060743
REFERENCE 2 (bases 1 to 82746)
AUTHORS Yu,T.W., Bai,L., Clade,D., Hoffmann,D., Toelzer,S., Trinh,K.Q.,
TITLE Xu,J., Moss,S.J., Leistner,E. and Floss,H.G.
JOURNALS Direct Submission
Submitted (28-NOV-2001) Department of Chemistry, University of
Washington, Box 351700, Seattle, WA 98195-1700, USA
FEATURES
Source Location/Qualifiers
1. 82746
/organism="Actinosynnema pretiosum subsp. auranticum"
/strain="ATCC 31565"
/sub_species="auranticum"
/db_xref="ATCC:31565"
/db_xref="taxon:42198"
misc_feature <1. 82653
/note="cluster I of the maytansinoid antitumor agent
ansamitocin biosynthetic gene cluster"
gene <1. 1678
/gene="asm37"
CDS <1. 1678
/gene="asm37"
/codon_start=2
/transl_table=1
/product="unknown"
/protein_id="AAM54115.1"
/db_xref="GI:21449383"
/translation="1VPLVVALATGVLAPASAAITSPVTSTANTVTEVSGT
PASFERYVDLDRSAATGYAOGIGADHLENGSLYRHSGGAMTQVATVPTASGV
ARFVERADLGEOSAPNDVYFOVEAPVETSAVTHTYTAGGSVAVOYPPSSANVP
PERGLHHTGDCDKADLSADLRRTGEGITLACVPTYLAEHRSSPTAOTLRDOR
QEDARVAGCKVLIRFAYTASESESDAPKREVLAMHITQIALPLRANADYVAVAGPY
GAWGEMTYTKRFGAGVVSATDQANKRANVDKLTALPADRAVOLRTPKFKRALYGG
STRIGHNDEGLADTDFTGYEDPATEPYLEATRTLPMGETCAVNPPTDPTAT
AELSRFMTYLTNDYQODVLRAMSAOGLAEVATKLGRTLRDSVFPPTTBSAALP
VRVLTINDGYAAVNORTVNLRENTTGAITRLPLTADPTWMLPGARAIQDAQIP
SLATGCVALLLELADPALPGRPREXSIOPTANGMDARTGLNDKQITTVS"
1825. 2142
gene /gene="asm36"
CDS 1825. 2142
/gene="asm36"
/codon_start=1
/transl_table=1
/product="unknown"
/protein_id="AAM54114.1"
/db_xref="GI:21449382"
/translation="MVTASAVLCARAGCGAPRPGATARMFCSYACKQAAVRLRKA
GSASHEHEDQLREALLALVARARTITEAPPMGAQATALVGLARVVELAAADAGV
SGC"
gene complement(2129. 3307)
CDS /gene="asm35"
complement(2129. 3307)
/gene="asm35"
/codon_start=1
/transl_table=1
/product="bicyclomycin resistant protein"
/protein_id="AAM54113.1"
/db_xref="GI:21449381"
/translation="MAVGKADSAVGMPLPALLIYALTGTTIIMSVSPFDIARELG
VGGPVOYLTJLNFGLCYGLCHGPLSDRGRRPVLGGSAVNLVGLGALAPTTG
LLTARFVGGTGAASAIVIVIIIGDRYRGPALRMYGIYVAAAMSLMTGCPILGATG
NRVAGMGNLIYACATLALSLLVFLPERTRARARITARSYLADYRRLGSRATL
AASLTPSLTNAIVFTIASPFLITTAAGLCVTSFALHLLVVASFPVPSFAFGIIP
LLGGPAGAVRNGLAAVYVYGLLLDGPVAVGSLALFCVGFVAVCVPPVDFDSMSY
FPFLAGAASSATMGLRALVLTLLTALAGALSTGPFAPAAVMAAGAVAAVLLSLSPR
"

gene complement(3450. 4064)
CDS /gene="asm34"
complement(3450. 4064)
/gene="asm34"
/codon_start=1
/transl_table=1
/product="transcriptional regulator"
/protein_id="AAM54112.1"
/db_xref="GI:21449380"
/translation="MARNPERRAALDAAYEVLAGECARGLTRANDQRCVAGTA
SNYFAGRAEILTGAGERYQRLLDQAPLGETPDGRDRSTELMHAILVDVSAFPG
FIALLELEAARNPELREVEVLARVARDLRNNAHNAASGLPGGTTVALMLMLNL
ILERTLPDLTLDHQRDVLTALVDRIIACQPEBAGSPFDESGR"
4151. 4723
gene /gene="asm33"
CDS 4151. 4723
/gene="asm33"
/codon_start=1
/transl_table=1
/product="dihydrofolate reductase"
/protein_id="AAM54111.1"
/db_xref="GI:21449379"
/translation="MRKLIVFIAASIDGVIAPEGCHDFYTPAEPMEHLRAEYPTT
PTHYRPLVGMPLDTPNRRYDVLWGRGSYQPALDAGVSSPAHLROYVSSRLDADP
AVALVRADPVLVLELKRKEDGLDVMLCGALDAGQLPEIDELVTKTYVVAAGGTPL
FRAGFAPARPRPVVRAFDHGGVTTYRPA"
complement(4854. 5057)
gene /gene="asm32"
CDS complement(4854. 5057)
/gene="asm32"
/codon_start=1
/transl_table=1
/product="unknown"
/protein_id="AAM54110.1"
/db_xref="GI:21449378"
/translation="MWITQEARIPARRSSGRSNARCAVWMPBPQPSATRADGPS
SANDSRASRGDRPRTSLNSA"
5415. 6461
gene /gene="asm31"
CDS 5415. 6461
/gene="asm31"
/codon_start=1
/transl_table=1
/product="RNA polymerase sigma factor"
/protein_id="AAM54109.1"
/db_xref="GI:21449377"
/translation="MWIGYSADGRRRRPVRWGVTEAFWTETDEDIKIATLRDAEA
ANADSTRAYLKAIGRVRLTADQEVELARVEGLVAEHLDELPPAARGSRPVELR
RVVRDGORAKDHLLEANRLVYSIAKRYTGMAILLVIOGNGLVAVAEKFDYAKG
EKFSYVATMWIROATRLRADOSRTIRLPVWVAVVTKLRIEERDLHRRNREATAEE
VAVELDYTPPEKVVLELROHARTPVSLDQITGAEATPSIGDLTADENAYAPRVSALL
SADLRVAVLETLTPREAGVVRMYRGPVGLTLDQEVGVGTVERKRLQIESRWAKL
QHPARSRGLEGYLD"
complement(6501. 9518)
gene /gene="asm30"
CDS complement(6501. 9518)
/gene="asm30"
/codon_start=1
/transl_table=1
/product="cytochrome P450"
/protein_id="AAM54108.1"
/db_xref="GI:21449376"
/translation="MVAITGTRIPGKPLPYGNLLDVTSLDLPDVDFLDRCHREHG
ITALTFAGQROVFASSHLELVARKCSDPSWGSAVHPALAEQVNDPAGDGLFARGDEPNM
GKAHRLMPAGPTAMRDRHPAMLDIAEQMLVRRRRGPPDRIDVADMDPRLDITLA
LCAFGARFNSFYRDRAPHFVDAMRVSLVEAGERLRLPGVPLFVGNQRDRIADTM
NRIDGIVAAARALPAGRPDCLERMLTCADPYTGELSARNTYOLAPFLILHET
TSGILSPVHRLIAPRERKRAQVGVGDRPAREDLARLDVIGOVLEBTLRLHP
TAPFALAPDEPRLGHAIGAGEPVLYMLPTLHRDRAVWDPVFPDERAPERMDE
ITPACAMPEFGGARACIGRPALDQEAIVLVLQKRDLAADDPHRLITIKOTLTLRP
DSLVARMPFRADRGAGATAYTALDQVHPATHRGTPLHVYGSNGSGEGLARTIG
DGAARGATSVAPLDADVRAIPASGPVIVSSSYNGAPPDNAAHFVWMLTDGPDLSG
VDVILVIGGNDWMSATYGVPTLLIDEMAAAGARRLRERATIDARADPFQGWERYEP
"

BACPAC Resources (<http://bacpac.med.bufileo.edu/orderingframe.htm>)
or from Resea ch Genetics ([infereesgen.com](http://www.tlgr.org/tcd/bac_ends/mouse/bac_end_intro.html)). BAC end page:
http://www.tlgr.org/tcd/bac_ends/mouse/bac_end_intro.html
Plate: 458 row: N column: 8
Seq primer: SP6
Class: BAC ends.

FEATURES
source
Location/Qualifiers
1. .407

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-458N8"
/clone_1lb="RPCI-23"
/sex="Female"
/lab_host="DH10B"

/note="Organ: Kidney/Brain; Vector: pBACe3.6; Site: 1;
EcoRI; Site: 2; EcoRI; Female C57BL/6J mouse kidney and/or
brain genomic DNA was isolated and partially digested
with a combination of EcoRI and EcoRI Methylase. Size
selected DNA was cloned into the pBACe3.6 vector at the
EcoRI sites. The ligation products were transformed into
DH10B electrocompetent cells (BRL Life Technologies)."

BASE COUNT
66 a 136 c 132 g 73 t

ORIGIN

Query Match 39.4%; Score 39.8; DB 17; Length 407;
Best Local Similarity 67.5%; Pred. No. 0.68;
Matches 56; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

OY 4 GCATGCCGCTCCGCGGCTTCATGCGCGGCTATCCTGACGACGCGCTCTGATGC 63

DB 22 GCGTGGCTCTCCCGCGGCGGCTTCAGCTGAGCGGCTTCCTGAGAGAGTCTGATGC 81

OY 64 AGCAACACCCCGCGGCTTCGCG 86

DB 82 ATCAACACCCCGCGGCTTCGCG 104

RESULT 2
BG048070 575 bp mRNA linear EST 25-JAN-2001
LOCUS OVL27.C06.g1.A002 Ovary 1 (OVL) Sorghum bicolor cDNA, mRNA
DEFINITION
Sequence.
ACCESSION BG048070
VERSION BG048070.1 GI:12498451
KEYWORDS EST.
SOURCE sorghum.
ORGANISM Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Sorghum.
1 (bases 1 to 575)
Cordonier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
,L.H.

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
An EST database from Sorghum: ovaries of varying immature stages
Unpublished (2000)
Contact: Cordonier-Pratt MM
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Km. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: mmpratt@uga.edu

Sequences have been trimmed to exclude polyA, vector and regions
below Phred quality 16. The threshold for highest quality sequence
is 20.

FEATURES
source
Seq primer: PolyTMix
High quality sequence start: 63
High quality sequence stop: 574
POLYA-No.
Location/Qualifiers
1. .575
/organism="Sorghum bicolor"

/db_xref="taxon:4558"
/clone_1lb="Ovary 1 (OVL)"
/note="Organ: Mix of ovaries of varying immature stages
from 8-week-old plants; Vector: pBluescript II from Lambda
zap II; Site: 1: XhoI; Site: 2: EcoRI; The library was made
from poly-A RNA in the cloning vector Lambda Zap II.
Clones to be sequenced were prepared by mass excision."

BASE COUNT
108 a 180 c 168 g 119 t

Query Match 35.8%; Score 36.2; DB 12; Length 575;
Best Local Similarity 60.8%; Pred. No. 6.1;
Matches 59; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

OY 2 GGGATGCCGCTCCGCGGCTTCATGCGCGGCTATCCTGACGACGCGGCTGCTGAT 61

DB 26 GCGGACGCGCGCCACGCGGCGAGGTGGCGCGCGCAGATCTGCTGCGGCCCAT 85

OY 62 GCAGCAACCCCGCGGCTTCGCGGCTCATGCGG 98

DB 86 CCGGACGCGCAAGCGAGGCTCGCGCGCAGCTTCGCG 122

RESULT 3
BM628936 487 bp mRNA linear EST 26-FEB-2002
LOCUS 17000687499324 A.Gam.ad.cDNA1 Anopheles gambiae cDNA clone
DEFINITION 17000449626935 5', mRNA sequence.
ACCESSION BM628936
VERSION BM628936.1 GI:18928447
KEYWORDS EST.
SOURCE African malaria mosquito.
ORGANISM Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
Anopheles.
1 (bases 1 to 487)
Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L., Charlab
,R., Collins,F.H., Venter,J.C. and Hoffman,S.L.
Celera Anopheles gambiae EST project
Unpublished (2002)
Contact: Holt R.A.
Celera Genomics
45 W. Gude Dr., Rockville, MD 20850, USA
Tel: 2404533151
Fax: 2404534580
Email: Holt@celera.com
Plate: NU010049V row: E column: 13
Seq primer: M13 Reverse.
Location/Qualifiers
1. .487
/organism="Anopheles gambiae"
/strain="RSP-ST (Reduced susc. to Permethrin - std.
chromosome)"
/db_xref="taxon:7165"
/clone="19600449626935"
/clone_1lb="A.Gam.ad.cDNA1"
/dev_stage="Adult"
/lab_host="DH10B"

/note="Vector: pSport1; Site: 1: SalI; Site: 2: NotI; Whole
adult mosquitoes (mixed sex) frozen on liquid nitrogen.
cDNA inserts >500 bp cloned directionally into pSport 1.
Not 1 site is 3'. Clones available through the Malaria
Research and Reference Reagent Resource Center
(www.malaria.mt4.org)."

FEATURES
source
Location/Qualifiers
1. .487

BASE COUNT
81 a 145 c 185 g 76 t

Query Match 34.1%; Score 34.4; DB 13; Length 487;
Best Local Similarity 60.9%; Pred. No. 18;
Matches 56; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

OY 9 CCGCTCCGCGGCTTCATGCGCGGCTATCCTGACGACGCGGCTGATGACGCAA 68

Db 100 CCGGTCGGGGGTGTCTATCTGGCCGCTACGCGGAGCGGCTGCTGGAGAACTCG 159
Oy 69 ACCCCCGCGCGGTTCGCGCGGTATGCGGCC 100
Db 160 ACCGAAGCGGACTTCGCGCGGTGCGGCC 191

RESULT 4
BM640213
LOCUS 668 bp mRNA linear EST 26-FEB-2002
DEFINITION 17000687678931 A.Gam.ad.cDNA1 Anopheles gambiae cDNA clone
19600449622135 5', mRNA sequence.
ACCESSION BM640213
VERSION BM640213.1 GI:18939724
KEYWORDS EST.
SOURCE African malaria mosquito.
ORGANISM Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
Anopheles.
REFERENCE 1 (bases 1 to 668)
AUTHORS Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L., Charlab
,R., Collins,F.H., Venter,J.C. and Hoffman,S.L.
Celera Anopheles gambiae EST project
JOURNAL Unpublished (2002)
COMMENT Contact: Holt R.A.
Celera Genomics
45 W. Gude Dr., Rockville, MD 20850, USA
Tel: 2404533151
Fax: 2404534580
Email: HoltRA@celera.com
Plate: NU01004B12 row: M column: 13
Seq primer: M13 Reverse.
Location/Qualifiers

FEATURES
source 1..668
/organism="Anopheles gambiae"
/strain="RSP-ST (Reduced susc. to Permethrin - std.
chromosome)"
/db_xref="taxon:7165"
/clone="19600449622135"
/clone_11b="A.Gam.ad.cDNA1"
/dev_stage="Adult"
/lab_host="DH10b"
/note="Vector: pSport1; Site_1: SalI; Site_2: NotI; Whole
adult mosquitoes (mixed sex) frozen on liquid nitrogen.
cDNA inserts >500 bp cloned directionally into pSport 1.
Not 1 site is 3'. Clones available through the Malaria
Research and Reference Reagent Resource Center
(www.malaria.mr4.org)."

BASE COUNT 115 a 196 c 253 g 104 t
ORIGIN

Query Match 34.1%; Score 34.4; DB 13; Length 668;
Best Local Similarity 60.9%; Pred. No. 18;
Matches 56; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

Oy 9 CCGCTCCGGCGGCTGTCATGCGCGGTATCTGAGACGCGGCTGTCATGAGCAA 68
Db 366 CCGGTCCGCGGTGTCTAATCTGCGCGGTACTGCGGAGCGGCTGTCGAGAACTCG 425
Oy 69 ACCCCCGCGGTTCCGCGCGGTCTATGCGGCC 100
Db 426 ACCGAAGCGGACTTCGCGCGGTGCGGCC 457

RESULT 5
BM630368 670 bp mRNA linear EST 26-FEB-2002
LOCUS 17000687501332 A.Gam.ad.cDNA1 Anopheles gambiae cDNA clone
DEFINITION 19600449618253 5', mRNA sequence.
ACCESSION BM630368
VERSION BM630368.1 GI:18929879

KEYWORDS EST.
SOURCE African malaria mosquito.
ORGANISM Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
Anopheles.
REFERENCE 1 (bases 1 to 670)
AUTHORS Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L., Charlab
,R., Collins,F.H., Venter,J.C. and Hoffman,S.L.
Celera Anopheles gambiae EST project
JOURNAL Unpublished (2002)
COMMENT Contact: Holt R.A.
Celera Genomics
45 W. Gude Dr., Rockville, MD 20850, USA
Tel: 2404533151
Fax: 2404534580
Email: HoltRA@celera.com
Plate: NU01004N8G row: K column: 19
Seq primer: M13 Reverse.
Location/Qualifiers

FEATURES
source 1..670
/organism="Anopheles gambiae"
/strain="RSP-ST (Reduced susc. to Permethrin - std.
chromosome)"
/db_xref="taxon:7165"
/clone="19600449618253"
/clone_11b="A.Gam.ad.cDNA1"
/dev_stage="Adult"
/lab_host="DH10b"
/note="Vector: pSport1; Site_1: SalI; Site_2: NotI; Whole
adult mosquitoes (mixed sex) frozen on liquid nitrogen.
cDNA inserts >500 bp cloned directionally into pSport 1.
Not 1 site is 3'. Clones available through the Malaria
Research and Reference Reagent Resource Center
(www.malaria.mr4.org)."

BASE COUNT 119 a 192 c 248 g 111 t
ORIGIN

Query Match 34.1%; Score 34.4; DB 13; Length 670;
Best Local Similarity 60.9%; Pred. No. 18;
Matches 56; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

Oy 9 CCGCTCCGGCGGCTGTCATGCGCGGTATCTGAGACGCGGCTGTCATGAGCAA 68
Db 455 CCGGTCCGCGGTGTCTAATCTGCGCGGTACTGCGGAGCGGCTGTCGAGAACTCG 514
Oy 69 ACCCCCGCGGTTCCGCGCGGTCTATGCGGCC 100
Db 515 ACCGAAGCGGACTTCGCGCGGTGCGGCC 546

RESULT 6
BM606844 694 bp mRNA linear EST 25-FEB-2002
LOCUS 17000687084364 A.Gam.ad.cDNA blood1 Anopheles gambiae cDNA clone
DEFINITION 19600449697069 5', mRNA sequence.
ACCESSION BM606844
VERSION BM606844.1 GI:18904948
KEYWORDS EST.
SOURCE African malaria mosquito.
ORGANISM Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
Anopheles.
REFERENCE 1 (bases 1 to 694)
AUTHORS Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L., Charlab
,R., Collins,F.H., Venter,J.C. and Hoffman,S.L.
Celera Anopheles gambiae EST project
JOURNAL Unpublished (2002)
COMMENT Contact: Holt R.A.
Celera Genomics
45 W. Gude Dr., Rockville, MD 20850, USA
Tel: 2404533151


```

/clone="IMAGF:5557893"
/clone_lib="NIH_MGC_72"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: skin; Vector: pCMV-SPOB6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: oligo dT
Average insert size 2 kb. Library constructed by Life
Technologies."

```

	Query Match	Similarity	Score	DB	Length
Best Local	59.9%	32.9%	33.2	14	1059
Matches	56	Conservative	0	Mismatches	38
				Indels	0
				Gaps	0

RESULT 14					
B1227386		1221 bp	mRNA	linear	EST 11-JUL-2001
LOCUS	B1227386				
DEFINITION	B1227386	1221 bp	mRNA	linear	EST 11-JUL-2001
	602935040351	NIH_MGC_8	Homo sapiens	CDNA clone IMAGE:5094528 5',	
	mRNA sequence.				

ACCESSION	BI227386	
VERSION	BI227386.1	GI:14680830
KEYWORDS	EST.	
SOURCE	human.	

ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
Homo sapiens					
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.					
1 (bases 1 to 1221)					
NIH-MGC http://mgc.nci.nih.gov/					
National Institutes of Health, Mammalian Gene Collection (MGC)					
unpublished (1999)					
Contact: Robert Strausberg, Ph.D.					

Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue procurement: Louis M. Straud, M.D., Ph.D.
cDNA library preparation: Ling Hong/Rubin Laboratory
cDNA library arrayed by: The I.M.A.G.E. Consortium (LIML)
DNA sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LIML at:
<http://image.llnl.gov>
plate: L1CM1863 row: f column: 01
high quality sequence start: 3
high quality sequence stop: 689.

FEATURES
source

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5094528"
/clone_1lb="N1H.MC.8"
/tissue_type="Burkitt Lymphoma"
/lab_host="DH10b (phage-resistant)"
/note="Organ: Lymph; Vector: pOTB1; site_1: XhoI; Site-2: EcoRI; cDNA made by oligo-dr priming. Directionally cloned into EcoRI/XhoI sites using the following 5'-adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using Zap-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

```

Query Match
Best Local

32.9%; Score 33.2; DB 13; Length 1221;
59.6%; Pred. No. 38;

	Matches	56;	Conservative	0;	Mismatches	38;	Indels	0;	Gaps	0;
QY	8	GGCGTCCGGGGGCGTCATCGACGGCGGTCCTCGACGACGGGCGTCATGCAGCA	67							
Db	175	GCCCGTGGCGGGCGCTTTCACCTGGCCCGTGGCTTTGAGAGATGGCTTCTCGAATACCA	234							
QY	68	AACCCCGCGCGCTTCGCGCGGTCATGCGCCG	101							
Db	235	GACCCGAGGTTCTTCCAGACGACTGGAGAGCC	268							

RESULT 15	
BG674746	
LOCUS	845 bp mRNA linear
DEFINITION	602620858r1 NCI-CGAP-Skn3 Homo sapiens CDNA clone IMAGE:4746207 5',
ACCESSION	BG674746
VERSION	BG674746.1 GI:13906142
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

1 (bases 1 to 845)
NIN-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC).
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.

Tissue Procurement: James Cleaver, M.D.
cDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA
Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLML0594 row: d column: 16
High quality sequence stop: 772.

FEATURES
source

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4746207"
/clone_1lb="NCI CGAP Skn3"
/lab_host="DH10B (r1 phage-resistant)"
/note="Organ: skin; Vector: pCW-SPOPT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: oligo dT
Average insert size 1.5kb. Library constructed by Life
Technologies. Note: this is a NCI CGAP Library."
BASE COUNT
158 a 243 c 297 g 147 t
ORIGIN

```

Query Mat

	Query Match	Match Similarity	Best Local	Score	DB	Length
Matches	57	Conservative	0	Mismatches	40	Indels
						Gaps
Qy	5	GATGGCGGCGCCGGCGGCGTTCATGCGGGCGGATTCCTGAGACGACGGGCGTGAATCA	64			
Db	108	GGTGGCCCGGTGGCGGCGCGTTCATCACTGGCGCTGGTCTTGGAGAGATGGCTTCTGGAGAA	167			
Qy	65	GCAAAACCCCGCGCGGTTCCGCGCGGTCAATGGCGCCC	101			
Db	168	CCAGACCCCGAGTTCCTTCACGAGAGTCTGCAAGCCC	204			

Search completed: November 6, 2002, 15:52:48
Job time : 1201.62 secs

APPLICANT: Sherman D H

GENERAL INFORMATION:

```

: APPLICANT: Liu, H.
: APPLICANT: Xue, Y.
: APPLICANT: Zhao, L.
: TITLE OF INVENTION: DNA encoding methymycin and pikromycin
: FILE REFERENCE: 600,438US1
: CURRENT APPLICATION NUMBER: US/09/861,289
: CURRENT FILING DATE: 2001-05-18
: PRIOR APPLICATION NUMBER: 09/105,537
: PRIOR FILING DATE: 1998-06-26
: NUMBER OF SEQ ID NOS: 43
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 5
: LENGTH: 36778
: TYPE: DNA
: ORGANISM: Streptomyces venezuelae
: US-09-861-289-5

```

Query Match	35.68;	Score 36;	DB 10;	Length 36778;
Best Local Similarity	62.08;	Pred. NO. 0.046;		
Matches 57; Conservative	0;	Mismatches 35;	Indels 0;	Gaps 0;

Qy 1 CGGGGATGCCCGCTCCGGGGCGGTGTCATCAGCCGCATCCCTGGACGACGGGGCTGCTGA 60
| | | | | | | | | | | | | | | | | |
Db 8592 CCGAGACGCCCTCACCGCGCTGTCACACACGCGCGCGCGTTCGACGACGGCATCGTTGG 8651

Qy	61	TGCAGCAACCCCGCGCGGTTCCGCGCGGTC	92
Db	8652	ACAGCGTTGACCGCGAGCAGTCCGCGCGGCC	8683

RESULT 3
US-09-861-289-1

```

sequence 1, Application US/05060205
Patent No. US20020110897A1
GENERAL INFORMATION:
APPLICANT: Sherman, D. H.
APPLICANT: Liu, H.
APPLICANT: Xue, Y.
APPLICANT: Zhao, L.
TITLE OF INVENTION: DNA encoding methymycin and pikromycin
FILE REFERENCE: 600,438US1
CURRENT APPLICATION NUMBER: US/09/861,289
CURRENT FILING DATE: 2001-05-18
PRIOR APPLICATION NUMBER: 09/105,537
PRIOR FILING DATE: 1998-06-26
NUMBER OF SEQ ID NOS: 43
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 1
LENGTH: 15872
TYPE: DNA
ORGANISM: Streptomyces venezuelae
US-09-861-289-1

```

Query Match	34.1%;	Score 34.4;	DB 10;	Length 15872;
Best Local Similarity	60.9%;	Pred. NO. 0.13;		
Matches	56;	Conservative	0;	Mismatches 36;
			Indels 0;	Gaps 0;

Oy 9 CCGCTCCGGGCGTTCATGCGGCCGTATCCTGGACGACGGGCTGCTGATGCAGCA 68
 | | | | | | | | | | | | | | | | | |
Db 7577 CCCTGACGGCCGTCTGACACGCGGGAGTCGTGACGACGACGCTGGAAGCGCTC 7638

Oy 69 ACCCCCCGGGGTTCGGCGGTCATGGCGCC 100
 || || || || || || || ||
 Db 7637 ACACCGGAACGGCTGGACGGCGTACTGCGCC 7668

RESULT 4
US-09-925-301-474

Patent No. US20020052308A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies

```

? FILE REFERENCE: PAT106
? CURRENT APPLICATION NUMBER: US/09/925,301
? CURRENT FILING DATE: 2001-08-10
? PRIOR APPLICATION NUMBER: PCT/US00/05882
? PRIOR FILING DATE: 2000-03-08
? PRIOR APPLICATION NUMBER: 60/124,270
? PRIOR FILING DATE: 1999-03-12
? NUMBER OF SEQ ID NOS: 1694
? SOFTWARE: PatentIn Ver. 2.0
? SEQ ID NO 474
? LENGTH: 3209
? TYPE: DNA
? ORGANISM: Homo sapiens
? FEATURE:
? NAME/KEY: misc feature
? LOCATION: (4227
? OTHER INFORMATION: n equals a,t,g, or c
? US-09-925-301-474

```

Query Match	31.3%	Score 31.6;	DB 10;	Length 3209;
Best Local Similarity	58.5%;	Pred. No. 0.72;		
Matches	55;	Conservative	0;	Mismatches 39;
			Indels	0;
			Gaps	0;

```
Oy      8 GCGGCTCCGGGGCGTGTTCAATGCAGGCCGGTATCCTGGACGACGGGCTGTGATGACACA 67
         ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    724 GCCCGTGGGGCGTTCACACTGGCGGTGTTGAGAGATGGCTTGCCTGGNGNAAACA 783
```

```

Oy  68  AACCCTGGCGGTTCCGGCGGTCATGGCGCC 101
      |||| | ||| | |||
Db  784  GACCCAGAGTCTTCAGGACGTCTGCAAGCC 817

```

RESULT 5
US-09-861-289-42

? Sequence 42, Application 05/03601269
 ? Patent No. US2002010897A1
 ? GENERAL INFORMATION:
 ? APPLICANT: Sherman, D.H.
 ? APPLICANT: Liu, H.
 ? APPLICANT: Xue, Y.
 ? APPLICANT: Zhao, L.
 ? TITLE OF INVENTION: DNA encoding methymycin and pikromycin
 ? FILE REFERENCE: 600,438US1
 ? CURRENT APPLICATION NUMBER: US/09/861,289
 ? CURRENT FILING DATE: 2001-05-18
 ? PRIOR APPLICATION NUMBER: 09/105,537
 ? PRIOR FILING DATE: 1998-06-26
 ? NUMBER OF SEQ ID NOS: 43
 ? SOFTWARE: FastSeq for Windows Version 3.0
 ? SEQ ID NO 42
 ? LENGTH: 846
 ? TYPE: DNA
 ? ORGANISM: Streptomyces venezuelae
 ? US-09-861-289-42

Query Match	30.9%	Score 31.2;	DB 10;	Length 846;
Best Local Similarity	57.0%;	Pred. No. 0.88;		
Matches 57; Conservative	0;	Mismatches 43;	Indels 0;	Gaps 0;

Qy 1 CCGGATGCCGCCTCCGGGGCTGTATTGGGGCGTACTGACAGCGGGCTGCTGA 60
 | | | | | | | | | | | | | | | | | | | | | |
Db 446 CCGAGATCCGGCGGCTCAGCGCACGACGAGGGTTCTTCACGACGACAGCTGCTGC 505

```

qy      61 TCGAGCAAAACCCCGCGCGGTTCCGCGCGGTCATGGCGCC 100
      || | ||||| | ||| | |||||
Db     506 GCGTGGTCTGCCCCGCGCTGCGCAGCAGACTACAAGCGGC 545

```

RESULT 6
US-09-861-289-3

; Sequence 3, Application US/09861285
; Patent No. US20020110897A1
; GENERAL INFORMATION:

APPLICANT: Sherman, D.H.
APPLICANT: Liu, H.
APPLICANT: Xue, Y.
APPLICANT: Zhao, L.
TITLE OF INVENTION: DNA encoding methymycin and pikromycin
FILE REFERENCE: 600,438051
CURRENT APPLICATION NUMBER: US/09/861,289
CURRENT FILING DATE: 2001-05-18
PRIOR APPLICATION NUMBER: 09/105,537
PRIOR FILING DATE: 1998-06-26
NUMBER OF SEQ ID NOS: 43
SOFTWARE: fastseq for Windows Version 3.0
SEQ ID NO 3
LENGTH: 13613
TYPE: DNA
ORGANISM: Streptomyces venezuelae
US-09-861-289-3

Query Match 30.9%; Score 31.2; DB 10; Length 13613;
Best Local Similarity 57.0%; Pred. No. 1;
Matches 57; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 1 CGGGATGCCCTCCGGCGCTTCATCGCGCGGTATCTGTGACGAGCGGCTGTGA 60
DB 409 CCGAGATCCGGCGGCTCACCGACCGAGCGGTTCTCTCAGGACGAGCTGTGCG 468
QY 61 TGCAGCAAAACCCCGCGGCTCCCGCGGTCATCGCGCC 100
DB 469 GCGTGCTGCTGCCCGCGGCTCGCAGCGACTACAGGCGCG 508

RESULT 7

US-09-815-242-4083
Sequence 4083, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlssen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: fastseq for Windows Version 4.0
SEQ ID NO 4083
LENGTH: 1008
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-815-242-4083

Query Match 28.3%; Score 28.6; DB 10; Length 1008;
Best Local Similarity 64.2%; Pred. No. 4.8;

Matches 43; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 31 CGGCGGTATTCCTCGACGACGCGGCTCTGATGACGAAACCCCGCGGTTCCGCGG 90
DB 521 CGGCCAGTTTCATGCGCGAGTGTGCTGTGACGATGTGACGAGTATTCGCGCG 580
QY 91 TCATGCG 97
DB 581 GCATGCG 587

RESULT 8

US-09-921-992-2
Sequence 2, Application US/09921992
Patent No. US20020069426A1
GENERAL INFORMATION:
APPLICANT: Boronai, Albert;
APPLICANT: Campos, Narciso;
APPLICANT: Rodriguez-Concepcion, Manuel;
APPLICANT: Rohmer, Michel;
APPLICANT: Seeman, Myriam;
APPLICANT: Valentin, Henry E.;
APPLICANT: Venkatesh, Tyamagondlu V.;
APPLICANT: Venkatesh, Mylavaram;
TITLE OF INVENTION: Methyl-D-Erythritol Phosphate Pathway Genes
FILE REFERENCE: 16516, 107/35-21(51897)US
CURRENT APPLICATION NUMBER: US/09/921,992
CURRENT FILING DATE: 2001-08-06
PRIOR APPLICATION NUMBER: US 60/223,483
PRIOR FILING DATE: 2000-08-07
NUMBER OF SEQ ID NOS: 85
SEQ ID NO 2
LENGTH: 33675
TYPE: DNA
ORGANISM: Oryza sativa
FEATURE:
NAME/KEY: CDS
LOCATION: (6924)..(7019), (7163)..(7269), (7344)..(7444), (7525)..(7634),
LOCATION: (7694)..(7813), (7923)..(8153), (8253)..(8369), (8515)..(8589),
LOCATION: (9012)..(9011), (9163)..(9225), (9328)..(9472), (9589)..(9730),
LOCATION: (9951)..(10028), (10134)..(10293), (10694)..(10798),
LOCATION: (11028)..(11129)
NAME/KEY: unsure
LOCATION: (1..33675)
OTHER INFORMATION: unsure at all n locations
US-09-921-992-2

Query Match 28.3%; Score 28.6; DB 10; Length 33675;
Best Local Similarity 61.3%; Pred. No. 5.6;
Matches 46; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 27 CATGCGCGCGGTATTCCTGACGACGCGGCTGTGATGACGAAACCCCGCGGTTCCGC 86
DB 5429 CAGCAGCAGCTGCGCAGCGGAGCGGCTGTGACGAGTCCACGACGCGCGCG 5488
QY 87 GCGGTATGCGCGCC 101
DB 5489 GCGGTGCGGCGCGC 5503

RESULT 9

US-09-815-242-7845
Sequence 7845, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlssen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard

;; TITLE OF INVENTION: Identification of Essential Genes in
;; FILE OF INVENTION: Prokaryotes
;; FILE REFERENCE: ELITRA.011A
;; CURRENT APPLICATION NUMBER: US/09/815,242
;; PRIOR FILING DATE: 2001-03-21
;; PRIOR APPLICATION NUMBER: 60/191,078
;; PRIOR FILING DATE: 2000-03-21
;; PRIOR APPLICATION NUMBER: 60/206,848
;; PRIOR FILING DATE: 2000-05-23
;; PRIOR APPLICATION NUMBER: 60/207,727
;; PRIOR FILING DATE: 2000-05-26
;; PRIOR APPLICATION NUMBER: 60/242,578
;; PRIOR FILING DATE: 2000-10-23
;; PRIOR APPLICATION NUMBER: 60/253,625
;; PRIOR FILING DATE: 2000-11-27
;; PRIOR APPLICATION NUMBER: 60/257,931
;; PRIOR FILING DATE: 2000-12-22
;; PRIOR APPLICATION NUMBER: 60/269,308
;; PRIOR FILING DATE: 2001-02-16
;; NUMBER OF SEQ ID NOS: 14110
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 7845
;; LENGTH: 1506
;; TYPE: DNA
;; ORGANISM: Pseudomonas aeruginosa
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: (1)...(1506)
US-09-815-242-7845

Query Match 28.1%; Score 28.4; DB 10; Length 1506;
Best Local Similarity 58.1%; Pred. No. 5.6;
Matches 50; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 13 TCCGCGCGCTTCATCGCGCGGCTCTCGACGACGCGGCTGCTATGCAAGAACCC 72
DB 569 TCCGCGCTTCCTCAGCGCGGCTTCCTCGAAGTGAGACCCCGCTGCGACCA 628
QY 73 CCGCGCGGTTCGCGCGGCTATGCGC 98
DB 629 TCCCGCGCGCGCGCGCGCGCAAGCCG 654

RESULT 10
US-09-867-550-303
; Sequence 303, Application US/09867550
; Patent No. US20020082206A1
; GENERAL INFORMATION:
; APPLICANT: Leach, Martin D.
; APPLICANT: Mehraban, Foad,
; APPLICANT: Conley, Pamela
; APPLICANT: Law, Debbie
; APPLICANT: Topper, James
; TITLE OF INVENTION: No. US20020082206A1a Polynucleotides from Atherogenic Cells and
; FILE OF INVENTION: Thereby
; FILE REFERENCE: 21402-013 (Cura-313)
; CURRENT APPLICATION NUMBER: US/09/867,550
; CURRENT FILING DATE: 2001-09-20
; PRIOR APPLICATION NUMBER: USSN 60/208,427
; PRIOR FILING DATE: 2000-05-30
; NUMBER OF SEQ ID NOS: 2125
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 303
; LENGTH: 246
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)..(2)
; OTHER INFORMATION: wherein any n is one of a or t or g or c
US-09-867-550-303

Query Match 27.7%; Score 28; DB 10; Length 246;

Best Local Similarity 56.5%; Pred. No. 6.7;
Matches 52; Conservative 0; Mismatches 40; Indels 0; Gaps 0;
QY 9 CCGCTCCGCGCGCTTCATGCGCGCGGCTATCTGACGACGCGCTGCTGATGACCA 68
DB 40 CCGACCGCGCATGTCGACCGCGCGGACATCTGCGCGCATGCGCTGAGATGCGC 99
QY 69 ACCCGCGCGGTTCCGCGCGGCTATGCGCGC 100
DB 100 ACCGAGACCGCTTCGCGGACGCGCATCTCTC 131

RESULT 11
US-09-815-242-7789
; Sequence 7789, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Traawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE OF INVENTION: Prokaryotes
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7789
; LENGTH: 1032
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1032)
US-09-815-242-7789

Query Match 27.5%; Score 27.8; DB 10; Length 1032;
Best Local Similarity 55.8%; Pred. No. 8.1;
Matches 53; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

QY 5 GATGCGCGCTCCGCGCGCTTCATGCGCGCGGCTATCTGACGACGCGCTGCTGATGCA 64
DB 432 GCTGCTGCTGCTCGCGCTGCGCGCATGCTCGCGCGCGCTGCGCGCATGCTGCTGCT 491
QY 65 GCAACCGCGCGCGGCTTCGCGCGGCTATGCGCGC 99
DB 492 GCGCGCGGCTCGCGGCTGCTGCTGCTGCTGCGC 526

RESULT 12
US-10-002-600-62
; Sequence 62, Application US/10002600
; Patent No. US20020137077A1
; GENERAL INFORMATION:

APPLICANT: Hopkins, Christopher M.
APPLICANT: Peterson, David P.
APPLICANT: Cocks, Benjamin G.
APPLICANT: Hawkins, Phillip R.
TITLE OF INVENTION: GENES REGULATED IN ACTIVATED T CELLS
FILE REFERENCE: PA-0042 US
CURRENT APPLICATION NUMBER: US/10/002,600
CURRENT FILING DATE: 2001-10-25
PRIOR APPLICATION NUMBER: 60/243,521
PRIOR FILING DATE: 2000-10-25
NUMBER OF SEQ ID NOS: 116
SOFTWARE: PERL Program
SEQ ID NO 62
LENGTH: 1699
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
OTHER INFORMATION: Template ID: 23849.1
US-10-002-600-62

Query Match 27.3%; Score 27.6; DB 12; Length 1699;
Best Local Similarity 60.8%; Pred. No. 9.4;
Matches 45; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

OY 4 GGATGCCGCTCCGGCGCTGTCATGCGCGGATCTGGAGAGCGGCTGATGC 63
DB 153 GGGTGGCTTCCGGACCGGACATGACGCGGCCCACTAAGACGACTGGCACCC 212

OY 64 AGCAACCCCGCGCG 77
DB 213 TACAACCTGGCGCG 226

RESULT 13
US-09-925-301-441/C
Sequence 441, Application US/09925301
Patent No. US20020052308A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: PA106
CURRENT APPLICATION NUMBER: US/09/925,301
CURRENT FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05882
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 1694
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 441
LENGTH: 1082
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1136)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (462)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (465)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (1074)
OTHER INFORMATION: n equals a,t,g, or c
US-09-925-301-441

Query Match 26.9%; Score 27.2; DB 10; Length 1082;
Best Local Similarity 55.2%; Pred. No. 12;
Matches 53; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

OY 4 GGATGCCGCTCCGGCGCTGTCATGCGCGGATCTGGAGAGCGGCTGATGC 63
DB 257 GGAGCAGACGGCGCGCGGCTGTGGTCCCGCAGAGGACAGCCGCTGTGATGC 198
OY 64 AGCAACCCCGCGCGGTTCCGCGGCTGATGCGC 99
DB 197 GCCACAGTCCCGGAGCTCCGAGCTGCCGCGGCGCTC 162

RESULT 14
US-09-768-826-27
Sequence 27, Application US/09768826
Patent No. US20020012966A1
GENERAL INFORMATION:
APPLICANT: Shi et al.
TITLE OF INVENTION: 18 human secreted proteins
FILE REFERENCE: PF512P1
CURRENT APPLICATION NUMBER: US/09/768,826
CURRENT FILING DATE: 2001-01-25
PRIOR APPLICATION NUMBER: PCT/US00/22350
PRIOR FILING DATE: 2000-08-15
PRIOR APPLICATION NUMBER: 60/148,759
PRIOR FILING DATE: 1999-08-16
NUMBER OF SEQ ID NOS: 61
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 27
LENGTH: 1558
TYPE: DNA
ORGANISM: Homo sapiens
US-09-768-826-27

Query Match 26.7%; Score 27; DB 10; Length 1558;
Best Local Similarity 56.0%; Pred. No. 14;
Matches 51; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

OY 4 GGATGCCGCTCCGGCGCTGTCATGCGCGGATCTGGAGAGCGGCTGATGC 63
DB 39 GGAGACCGCGCGCGGAGACGAGTGTGTGGGGGAGACTGTGCTGTCTGATC 98

OY 64 AGCAACCCCGCGCGGTTCCGCGGCTCAT 94
DB 99 CGCCCCCACCCTCTCTGCACTGCGCT 129

RESULT 15
US-09-944-413-77
Sequence 77, Application US/09944413
Patent No. US20020156004A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin
APPLICANT: Botstein, David
APPLICANT: Baton, Dan
APPLICANT: Ferrara, Napoleone
APPLICANT: Filvaroff, Ellen
APPLICANT: Gerritsen, Mary
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul
APPLICANT: Grimaldi, Christopher
APPLICANT: Gurney, Austin
APPLICANT: Hillan, Kenneth
APPLICANT: Kijavini, Ivar
APPLICANT: Napier, Mary
APPLICANT: Roy, Margaret
APPLICANT: Tumas, Daniel
APPLICANT: Wood, William
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
ACIDS ENCODING THE SAME
FILE REFERENCE: P254BP1C1
CURRENT APPLICATION NUMBER: US/09/944,413
CURRENT FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: 09/866,028
PRIOR FILING DATE: 2001-05-25
PRIOR APPLICATION NUMBER: 60/067,411

; Sequence 1, Application US/0

```
; Patent No. 6121029
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/335,409
; PRIORITY FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
; US-09-335-409-1
```

```
Query Match          98.4%; Score 99.4; DB 3; Length 68750;
Best Local Similarity 99.0%; Pred. No. 9e-19;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 CGGGATGCCGCTCCGCGGCTGTCATGCGCGGATCTGAGACGAGGCGTCTGA 60
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 42552 CGGGATGCCGCTCCGCGGCTGTCATGCGCGGATCTGAGACGAGGCGTCTGA 42611
```

```
QY 61 TGCAGCAAAACCCCGCGGCTTCGCGGCTCATGAGCGCC 101
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 42612 TGCAGCAAAACCCCGCGGCTTCGCGGCTCATGAGCGCC 42652
```

```
RESULT 3
; US-09-568-102-1
; Sequence 1, Application US/09568102
; Patent No. 6346404
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,102
; PRIORITY FILING DATE: 2000-05-10
; CURRENT APPLICATION NUMBER: 09/335,409
; PRIORITY FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
; US-09-568-102-1
```

```
Query Match          98.4%; Score 99.4; DB 4; Length 68750;
Best Local Similarity 99.0%; Pred. No. 9e-19;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 CGGGATGCCGCTCCGCGGCTGTCATGCGCGGATCTGAGACGAGGCGTCTGA 60
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 42552 CGGGATGCCGCTCCGCGGCTGTCATGCGCGGATCTGAGACGAGGCGTCTGA 42611
```

```
QY 61 TGCAGCAAAACCCCGCGGCTTCGCGGCTCATGAGCGCC 101
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 42612 TGCAGCAAAACCCCGCGGCTTCGCGGCTCATGAGCGCC 42652
```

```
RESULT 4
US-09-567-969-1
```

```
; Sequence 1, Application US/09567969
; Patent No. 6355457
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/567,969
; PRIORITY FILING DATE: 2000-05-10
; CURRENT APPLICATION NUMBER: 09/335,409
; PRIORITY FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
; US-09-567-969-1
```

```
Query Match          98.4%; Score 99.4; DB 4; Length 68750;
Best Local Similarity 99.0%; Pred. No. 9e-19;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 CGGGATGCCGCTCCGCGGCTGTCATGCGCGGATCTGAGACGAGGCGTCTGA 60
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 42552 CGGGATGCCGCTCCGCGGCTGTCATGCGCGGATCTGAGACGAGGCGTCTGA 42611
```

```
QY 61 TGCAGCAAAACCCCGCGGCTTCGCGGCTCATGAGCGCC 101
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 42612 TGCAGCAAAACCCCGCGGCTTCGCGGCTCATGAGCGCC 42652
```

```
RESULT 5
; US-09-568-480-1
; Sequence 1, Application US/09568480
; Patent No. 6355458
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,480
; PRIORITY FILING DATE: 2000-05-10
; CURRENT APPLICATION NUMBER: 09/335,409
; PRIORITY FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
; US-09-568-480-1
```

```
Query Match          98.4%; Score 99.4; DB 4; Length 68750;
Best Local Similarity 99.0%; Pred. No. 9e-19;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 CGGGATGCCGCTCCGCGGCTGTCATGCGCGGATCTGAGACGAGGCGTCTGA 60
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 42552 CGGGATGCCGCTCCGCGGCTGTCATGCGCGGATCTGAGACGAGGCGTCTGA 42611
```

```
QY 61 TGCAGCAAAACCCCGCGGCTTCGCGGCTCATGAGCGCC 101
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 42612 TGCAGCAAAACCCCGCGGCTTCGCGGCTCATGAGCGCC 42652
```

```
RESULT 6
US-09-568-486-1
; Sequence 1, Application US/09568486
; Patent No. 635459
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,486
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-486-1

Query Match          98.4%; Score 99.4; DB 4; Length 68750;
Best Local Similarity 99.0%; Pred. No. 9e-19;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGGGATGCGGCTCCGCGGCGTTCATGCGCGGATCTGAGCAGCGGCTGCTGA 60
Db 42552 CGGGATGCGGCTCCGCGGCGTTCATGCGCGGATCTGAGCAGCGGCTGCTGA 42611

QY 61 TGCAGCAAAACCCCGCGGCTTCGCGCGGTCATGCGGCC 101
Db 42612 TGCAGCAAAACCCCGCGGCTTCGCGCGGTCATGCGGCC 42652

RESULT 7
US-09-568-472-1
; Sequence 1, Application US/09568472
; Patent No. 6358719
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,472
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-472-1

Query Match          98.4%; Score 99.4; DB 4; Length 68750;
Best Local Similarity 99.0%; Pred. No. 9e-19;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGGGATGCGGCTCCGCGGCGTTCATGCGCGGATCTGAGCAGCGGCTGCTGA 60
Db 42552 CGGGATGCGGCTCCGCGGCGTTCATGCGCGGATCTGAGCAGCGGCTGCTGA 42611

QY 61 TGCAGCAAAACCCCGCGGCTTCGCGCGGTCATGCGGCC 101
Db 42612 TGCAGCAAAACCCCGCGGCTTCGCGCGGTCATGCGGCC 42652

RESULT 8
US-09-567-899-1
; Sequence 1, Application US/09567899
; Patent No. 6383787
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/567,899
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-567-899-1

Query Match          98.4%; Score 99.4; DB 4; Length 68750;
Best Local Similarity 99.0%; Pred. No. 9e-19;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGGGATGCGGCTCCGCGGCGTTCATGCGCGGATCTGAGCAGCGGCTGCTGA 60
Db 42552 CGGGATGCGGCTCCGCGGCGTTCATGCGCGGATCTGAGCAGCGGCTGCTGA 42611

QY 61 TGCAGCAAAACCCCGCGGCTTCGCGCGGTCATGCGGCC 101
Db 42612 TGCAGCAAAACCCCGCGGCTTCGCGCGGTCATGCGGCC 42652

RESULT 9
US-09-036-987A-1
; Sequence 1, Application US/09036987A
; Patent No. 6143526
; GENERAL INFORMATION:
; APPLICANT: Baltz, Richard H.
; APPLICANT: Broughton, Mary C.
; APPLICANT: Crawford, Kathryn P.
; APPLICANT: Madduri, Krishnamurthy
; APPLICANT: Merlo, Donald J.
; APPLICANT: Treadway, Patli J.
; APPLICANT: Turner, Jan R.
; APPLICANT: Waldron, Clive
; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
; TITLE OF INVENTION: Production
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dow Agrosciences LLC Patent Department
; STREET: 9330 Zionsville Road
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: USA
; ZIP: 46268
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/036,987A
; FILING DATE: 09-MAR-1998
```

CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Stuart, Donald R
REGISTRATION NUMBER: 28,479
REFERENCE/DOCKET NUMBER: 50,608
TELECOMMUNICATION INFORMATION:
TELEPHONE: (317)337-4816
TELEFAX: (317)337-4847
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 80161 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-09-036-987A-1

Query Match 44.6%; Score 45; DB 3; Length 80161;
Best Local Similarity 65.3%; Pred. No. 0.00047;
Matches 66; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

QY 1 CGGGATGCCGCTCCGGCGGCTGTCATGCGCGGATCTCGACGACGGGCTGCTGA 60
DB 74831 CGGAAACCGCTGCGGGGTGTGTGACGCCCGCGGTGCTGTGATGACGCTGTCTGA 74890

QY 61 TGCAGCAACCCCGCGCGGCTTCGCGCGGTCTCATGGCGCCC 101
DB 74891 TGTGATGTGCGCGGAGCGCTTGACGCGGCTGTGGCGCCC 74931

RESULT 10
US-09-370-700-1
Sequence 1, Application US/09370700
Patent No. 6274350
GENERAL INFORMATION:
APPLICANT: Baltz, Richard H
APPLICANT: Broughton, Mary C
APPLICANT: Crawford, Kathryn P
APPLICANT: Madduri, Krishnamurthy
APPLICANT: Treadway, Patl J
APPLICANT: Turner, Jan R
APPLICANT: Waldron, Clive
TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
FILE REFERENCE: 50489 DIV1
CURRENT APPLICATION NUMBER: US/09/370,700
CURRENT FILING DATE: 1999-08-09
EARLIER APPLICATION NUMBER: US 09/36987
EARLIER FILING DATE: 1998-03-09
NUMBER OF SEQ ID NOS: 39
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 1
LENGTH: 80161
TYPE: DNA
ORGANISM: Saccharopolyspora spinosa
US-09-370-700-1

Query Match 44.6%; Score 45; DB 4; Length 80161;
Best Local Similarity 65.3%; Pred. No. 0.00047;
Matches 66; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

QY 1 CGGGATGCCGCTCCGGCGGCTGTCATGCGCGGATCTCGACGACGGGCTGCTGA 60
DB 74831 CGGAAACCGCTGCGGGGTGTGTGACGCCCGCGGTGCTGTGATGACGCTGTCTGA 74890

QY 61 TGCAGCAACCCCGCGCGGCTTCGCGCGGTCTCATGGCGCCC 101
DB 74891 TGTGATGTGCGCGGAGCGCTTGACGCGGCTGTGGCGCCC 74931

RESULT 11
US-08-258-261B-6
Sequence 6, Application US/08258261B
Patent No. 5639949

GENERAL INFORMATION:
APPLICANT: Schnupp, Thomas
APPLICANT: Ligon, James M.
APPLICANT: Beck, James Joseph
APPLICANT: Hill, Dwight Steven
APPLICANT: Ryals, John Andrew
APPLICANT: Gaffney, Thomas Deane
APPLICANT: Lam, Stephen Ting
APPLICANT: Hammer, Phillip E.
APPLICANT: Uknes, Scott Joseph
TITLE OF INVENTION: Genes for the synthesis of
antipathogenic substances
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ciba-Geigy Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: NY
COUNTRY: USA
ZIP: 10532

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/258,261B
FILING DATE: 08-JUN-1994
CLASSIFICATION: 800

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/457,205
FILING DATE: 01-JUN-1995

ATTORNEY/AGENT INFORMATION:

NAME: Elmer, James Scott
REGISTRATION NUMBER: 36,129

REFERENCE/DOCKET NUMBER: CGC 1506/CIP3

TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8614
TELEFAX: 919-541-8689

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:
LENGTH: 28958 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOHETICAL: NO
ANTI-SENSE: NO
US-08-258-261B-6

Query Match 39.8%; Score 40.2; DB 1; Length 28958;
Best Local Similarity 64.5%; Pred. No. 0.0089;
Matches 60; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 9 CGGCTCCGGCGCTGCTTCATGCGCGCGGTATCTGACGACGGGCTGTGACGCAA 68
DB 2492 CCGCTACGCGCGCTGTCGACGCCCGCGGCGCTTGACGATGGCTGTACGACGANG 2551
QY 69 ACCCGCGCGGCTTCGCGCGGTCATGCGCGCCC 101
DB 2552 AGCCCGAGCGCATGACGCGCTTTGCTCCC 2584

RESULT 12

US-08-456-837-6
Sequence 6, Application US/08456837
Patent No. 5643774

GENERAL INFORMATION:

APPLICANT: Schnupp, Thomas
APPLICANT: Ligon, James M.
APPLICANT: Beck, James Joseph
APPLICANT: Hill, Dwight Steven
APPLICANT: Ryals, John Andrew

```

: APPLICANT: Gaffney, Thomas Deane
: APPLICANT: Lam, Stephen Ting
: APPLICANT: Hammer, Phillip E.
: APPLICANT: Uknes, Scott Joseph
: TITLE OF INVENTION: Genes for the synthesis of
: TITLE OF INVENTION: antipathogenic substances
: NUMBER OF SEQUENCES: 22
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Ciba-Geigy Corporation
: STREET: 7 Skyline Drive
: CITY: Hawthorne
: STATE: NY
: COUNTRY: USA
: ZIP: 10532
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/456,837
: FILING DATE: 01-JUN-1995
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/457,205
: FILING DATE: 01-JUN-1995
: APPLICATION NUMBER: 08/258,261
: FILING DATE: 08-Jun-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Elmer, James Scott
: REGISTRATION NUMBER: 36,129
: REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 919-541-8614
: TELEFAX: 919-541-8689
: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 28958 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: HYPOTHEICAL: NO
: ANTI-SENSE: NO
: US-08-456-837-6

Query Match          39.8%; Score 40.2; DB 1; Length 28958;
Best Local Similarity 64.5%; Pred. No. 0.0089;
Matches 60; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 9 CCGCTCCCGCGGCGTTCATGCGCGCGGTATCCTGGACGACGCGGCTGCTGATGCAGCAA 68
DB 2492 CCGCTCAGCGCGCGTGTGACGCGCGCGCGCTTGACGATGGCTGATCAGCGAATG 2551
QY 69 ACCCCCGCGGCTTCCGCGCGGTATGCGGCC 101
DB 2552 AGCCCGAGCGCATGACCGCGCTTGTCTCC 2584

RESULT 13
US-08-457-342-6
: Sequence 6, Application US/08457342
: Patent No. 5662898
: GENERAL INFORMATION:
: APPLICANT: Schnupp, Thomas
: APPLICANT: Ligon, James M.
: APPLICANT: Beck, James Joseph
: APPLICANT: Hill, Dwight Steven
: APPLICANT: Ryals, John Andrew
: APPLICANT: Gaffney, Thomas Deane
: APPLICANT: Lam, Stephen Ting
: APPLICANT: Hammer, Phillip E.
: APPLICANT: Uknes, Scott Joseph

```

```

: TITLE OF INVENTION: Genes for the synthesis of
: TITLE OF INVENTION: antipathogenic substances
: NUMBER OF SEQUENCES: 22
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Ciba-Geigy Corporation
: STREET: 7 Skyline Drive
: CITY: Hawthorne
: STATE: NY
: COUNTRY: USA
: ZIP: 10532
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/457,342
: FILING DATE: 01-JUN-1995
: CLASSIFICATION: 424
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/457,205
: FILING DATE: 01-JUN-1995
: APPLICATION NUMBER: 08/258,261
: FILING DATE: 08-Jun-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Elmer, James Scott
: REGISTRATION NUMBER: 36,129
: REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 919-541-8614
: TELEFAX: 919-541-8689
: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 28958 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: HYPOTHEICAL: NO
: ANTI-SENSE: NO
: US-08-457-342-6

Query Match          39.8%; Score 40.2; DB 1; Length 28958;
Best Local Similarity 64.5%; Pred. No. 0.0089;
Matches 60; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 9 CCGCTCCCGCGGCGTTCATGCGCGCGGTATCCTGGACGACGCGGCTGCTGATGCAGCAA 68
DB 2492 CCGCTCAGCGCGCGTGTGACGCGCGCGCGCTTGACGATGGCTGATCAGCGAATG 2551
QY 69 ACCCCCGCGGCTTCCGCGCGGTATGCGGCC 101
DB 2552 AGCCCGAGCGCATGACCGCGCTTGTCTCC 2584

RESULT 14
US-08-457-646A-6
: Sequence 6, Application US/08457646A
: Patent No. 5679560
: GENERAL INFORMATION:
: APPLICANT: Schnupp, Thomas
: APPLICANT: Ligon, James M.
: APPLICANT: Beck, James Joseph
: APPLICANT: Hill, Dwight Steven
: APPLICANT: Ryals, John Andrew
: APPLICANT: Gaffney, Thomas Deane
: APPLICANT: Lam, Stephen Ting
: APPLICANT: Hammer, Phillip E.
: APPLICANT: Uknes, Scott Joseph
: TITLE OF INVENTION: Genes for the synthesis of
: TITLE OF INVENTION: antipathogenic substances
: NUMBER OF SEQUENCES: 22
: CORRESPONDENCE ADDRESS:

```

```

: ADDRESSEE: Ciba-Gelgy Corporation
: STREET: 7 Skyline Drive
: CITY: Hawthorne
: STATE: NY
: COUNTRY: USA
: ZIP: 10532
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/457,646A
: FILING DATE: 01-JUN-1995
: CLASSIFICATION: 530
: PRIORITY APPLICATION DATA:
: APPLICATION NUMBER: US 08/457,205
: FILING DATE: 01-JUN-1995
: APPLICATION NUMBER: 08/258,261
: FILING DATE: 08-Jun-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Elmer, James Scott
: REGISTRATION NUMBER: 36,129
: REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 919-541-8614
: TELEFAX: 919-541-8689
: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 28958 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: HYPOTHEICAL: NO
: ANTI-SENSE: NO
:
: US-08-457-646A-6
:
: Query Match 39.8%; Score 40.2; DB 1; Length 28958;
: Best Local Similarity 64.5%; Pred. No. 0.0089;
: Matches 60; Conservative 0; Mismatches 33; Indels 0; Gaps 0;
:
: QY 9 CCGCTCCGGCGCTGCTTCATGCGCGCGGTATCTGGACGACGCGGCTGTGATGCAGCAA 68
: 11111 11111 11111 11111 11111 11111 11111 11111 11111 11111
:
: Db 2492 CCGCTCAGCGCGCTGCTGACGCGCGCGGCGGCGCTTGACGATGGCTGATCAGCAGATG 2551
:
: QY 69 ACCCGCGCGGCTTCGCGCGGTCATGCGCGCC 101
: 11111 11111 11111 11111 11111 11111 11111 11111 11111 11111
:
: Db 2552 AGCCCGAGCGCATGCGACGCGCTTTGCTGCC 2584
:
: RESULT 15
: US-08-458-076A-6
: Sequence 6, Application US/08458076A
: Patent No. 5698425
: GENERAL INFORMATION:
: APPLICANT: Schupp, Thomas
: APPLICANT: Ligon, James M.
: APPLICANT: Beck, James Joseph
: APPLICANT: Hill, Dwight Steven
: APPLICANT: Ryals, John Andrew
: APPLICANT: Gaffney, Thomas Deane
: APPLICANT: Lam, Stephen Ting
: APPLICANT: Hammer, Phillip E.
: APPLICANT: Uknes, Scott Joseph
: TITLE OF INVENTION: Genes for the synthesis of
: TITLE OF INVENTION: antipathogenic substances
: NUMBER OF SEQUENCES: 22
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Ciba-Gelgy Corporation
: STREET: 7 Skyline Drive
: CITY: Hawthorne
: STATE: NY
```

```

: COUNTRY: USA
: ZIP: 10532
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/458,076A
: FILING DATE: 01-JUN-1995
: CLASSIFICATION: 435
: PRIORITY APPLICATION DATA:
: APPLICATION NUMBER: US 08/457,205
: FILING DATE: 01-JUN-1995
: APPLICATION NUMBER: 08/258,261
: FILING DATE: 08-Jun-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Elmer, James Scott
: REGISTRATION NUMBER: 36,129
: REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 919-541-8614
: TELEFAX: 919-541-8689
: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 28958 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: HYPOTHEICAL: NO
: ANTI-SENSE: NO
:
: US-08-458-076A-6
:
: Query Match 39.8%; Score 40.2; DB 1; Length 28958;
: Best Local Similarity 64.5%; Pred. No. 0.0089;
: Matches 60; Conservative 0; Mismatches 33; Indels 0; Gaps 0;
:
: QY 9 CCGCTCCGGCGCTGCTTCATGCGCGCGGTATCTGGACGACGCGGCTGTGATGCAGCAA 68
: 11111 11111 11111 11111 11111 11111 11111 11111 11111 11111
:
: Db 2492 CCGCTCAGCGCGCTGCTGACGCGCGGCGGCGCTTGACGATGGCTGATCAGCAGATG 2551
:
: QY 69 ACCCGCGCGGCTTCGCGCGGTCATGCGCGCC 101
: 11111 11111 11111 11111 11111 11111 11111 11111 11111 11111
:
: Db 2552 AGCCCGAGCGCATGCGACGCGCTTTGCTGCC 2584
:
: Search completed: November 6, 2002, 17:29:13
: Job time : 211.625 secs
```


PR 13-APR-2001; 2001US-269020P.
XX
PA (KOSA-) KOSAN BIOSCIENCES INC.
XX
XX Arslanian RL, Ashley G, Frykman S, Julien B, Katz L, Khosla C;
PI Lau J, Licardi PJ, Regentin R, Santl D, Tang L;
XX
XX WPI; 2002-075167/10.
DR P-PSDB; ABB07167.
XX
PT Recombinant host cells useful for producing polyketides e.g. epothilone
PT or its derivatives, comprises a recombinant expression vector encoding
PS a heterologous polyketide synthase gene -
XX
XX Example 11; Page 164-165; 221pp; English.
XX
CC The invention provides a recombinant host cell, of the suborder
CC Cystobacterineae comprising a recombinant expression vector encoding a
CC heterologous polyketide synthase (PKS) gene and produces a polyketide
CC synthesized by the PKS enzyme encoded on the vector. An epothilone
CC derivative of a specified formula can be produced by culturing the host
CC cell with a diketide equivalent compound of a specified formula. The host
CC cells produces polyketides at high levels and are used in the production
CC of not only epothilones, including new epothilone derivatives, but also
CC other polyketides. Methods of purifying the epothilone derivatives are
CC also useful for treating cancer, hyperproliferative diseases and
CC conditions such as psoriasis, inflammation, sarcomas, neoplasms,
CC lymphomas, multiple sclerosis, rheumatoid arthritis, atherosclerosis and
CC /or restenosis. It improves polyketide production in any organism and
CC also for production of products of recombinant PKS genes and modification
CC enzymes. The present sequence represents the nucleotide sequence of the
CC KR domain of extender module 6 of the epothilone PKS. Inactivation of
CC this domain results in a novel PKS capable of producing a 9-keto-
CC epothilone analogue.
XX
SQ Sequence 552 BP; 70 A; 176 C; 214 G; 92 T; 0 other;
Query Match 100.0%; Score 101; DB 24; Length 552;
Best Local Similarity 100.0%; Pred. No. 6.3e-18;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CGGGGATGCCGTCGCCGCGGCTTCATCGCGCGGTATCTGACGACGCGGCTCTGA 60
DB 236 CGGGGATGCCGTCGCCGCGGCTTCATCGCGCGGTATCTGACGACGCGGCTCTGA 295
QY 61 TGCAGCAAAACCCCGCGCGGTCCTCCGCGGTCATGCGGCC 101
DB 296 TGCAGCAAAACCCCGCGCGGTCCTCCGCGGTCATGCGGCC 336
RESULT 2
ID AA168064 standard; DNA; 552 BP.
XX
XX AA168064:
XX
XX 13-MAR-2002 (first entry)
XX
XX 9-keto-epothilone PKS inactive KR domain nucleotide sequence.
DE
XX
XX Cystobacterineae; recombinant; polyketide synthase; PKS; polyketide;
KW epothilone; p11A gene; cytostatic; antiproliferic; antiarthritic;
KW antilactarisclosteric; antiinflammatory; neuroprotective; vasotropic; ds.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT CDS 1..552
FT /tag= a
FT /note= "mutated KR domain of extender module 6"
XX
XX WO200183800-A2.
XX
XX

PD 08-NOV-2001.
XX
XX
PF 26-APR-2001; 2001WO-US13793.
XX
XX
PR 28-APR-2000; 2000US-0560367.
PR 14-SEP-2000; 2000US-232696P.
PR 21-DEC-2000; 2000US-257517P.
PR 03-APR-2001; 2001US-0825856.
PR 03-APR-2001; 2001US-0825876.
PR 13-APR-2001; 2001US-269020P.
XX
XX (KOSA-) KOSAN BIOSCIENCES INC.
XX
XX Arslanian RL, Ashley G, Frykman S, Julien B, Katz L, Khosla C;
PI Lau J, Licardi PJ, Regentin R, Santl D, Tang L;
XX
XX WPI; 2002-075167/10.
DR P-PSDB; ABB07168.
XX
XX
PT Recombinant host cells useful for producing polyketides e.g. epothilone
PT or its derivatives, comprises a recombinant expression vector encoding
PS a heterologous polyketide synthase gene -
XX
XX Example 11; Page 165-166; 221pp; English.
XX
XX
CC The invention provides a recombinant host cell, of the suborder
CC Cystobacterineae comprising a recombinant expression vector encoding a
CC heterologous polyketide synthase (PKS) gene and produces a polyketide
CC synthesized by the PKS enzyme encoded on the vector. An epothilone
CC derivative of a specified formula can be produced by culturing the host
CC cell with a diketide equivalent compound of a specified formula. The host
CC cells produces polyketides at high levels and are used in the production
CC of not only epothilones, including new epothilone derivatives, but also
CC other polyketides. Methods of purifying the epothilone derivatives are
CC also useful for treating cancer, hyperproliferative diseases and
CC conditions such as psoriasis, inflammation, sarcomas, neoplasms,
CC lymphomas, multiple sclerosis, rheumatoid arthritis, atherosclerosis and
CC /or restenosis. It improves polyketide production in any organism and
CC also for production of products of recombinant PKS genes and modification
CC enzymes. The present sequence represents the nucleotide sequence of a
CC mutated and inactive KR domain of extender module 6 of the novel 9-keto-
CC epothilone PKS of the present invention.
XX
SQ Sequence 552 BP; 70 A; 177 C; 212 G; 93 T; 0 other;
Query Match 100.0%; Score 101; DB 24; Length 552;
Best Local Similarity 100.0%; Pred. No. 6.3e-18;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CGGGGATGCCGTCGCCGCGGCTTCATCGCGCGGTATCTGACGACGCGGCTCTGA 60
DB 236 CGGGGATGCCGTCGCCGCGGCTTCATCGCGCGGTATCTGACGACGCGGCTCTGA 295
QY 61 TGCAGCAAAACCCCGCGCGGTCCTCCGCGGTCATGCGGCC 101
DB 296 TGCAGCAAAACCCCGCGCGGTCCTCCGCGGTCATGCGGCC 336
RESULT 3
ID AAA29349 standard; DNA; 71989 BP.
XX
XX AAA29349:
XX
XX 12-SEP-2000 (first entry)
XX
XX Sorangium cellulosum epothilone polyketide synthase operon genomic DNA.
DE
XX
XX Epothilone; polyketide synthase; epOA; epOB; epOC; epOD; epOE; epOF;
KW epOL; epOK; P450 epoxidase; ORFV; ORFP; promoter; enhancer; anti-fungal;
KW tubulin polymerization assay; anti-tumour; cytostatic; ds.
XX
XX Sorangium cellulosum.
OS

XX	Key	Location/Qualifiers	FT	FT	misc_RNA	/note= "encodes adenylation domain A7 of the NRPS module"
FT	CDS	3..992	FT	FT	9090..9179	/**tag= t
FT		/**tag= a	FT	FT	/note= "encodes adenylation domain A8 of the NRPS module"	
FT		/label= ORF_A	FT	FT	9183..9992	/**tag= u
FT		/product= transposase	FT	FT	/note= "encodes oxidation region for forming thiazole"	
FT		/note= "not part of the PKS"	FT	FT	10121..10138	/**tag= v
FT	CDS	989..1501	FT	FT	/note= "encodes adenylation domain A10 of the NRPS module"	
FT		/**tag= b	FT	FT	10261..10306	/**tag= w
FT		/label= ORF_B	FT	FT	/note= "encodes thiolation domain (PCP) of the NRPS module"	
FT		/product= transposase	FT	FT	10639..16137	/**tag= x
FT		/note= "not part of the PKS"	FT	FT	/label= epoc_gene	
FT	CDS	1998..6263	FT	FT	/note= "encodes module 2"	
FT		/**tag= c	FT	FT	10654..12033	/**tag= y
FT		/label= epoc_gene	FT	FT	/note= "encodes KS2, the KS domain of module 2"	
FT	misc_RNA	/note= "encodes the loading domain"	FT	FT	12250..13287	/**tag= z
FT		2031..3548	FT	FT	/note= "encodes AT2, the AT domain of module 2"	
FT		/**tag= d	FT	FT	13327..13899	/note= "encodes dehydratase (DH) 2, the DH domain of module 2"
FT		/label= "encodes ketide synthase (KS-Y) of the loading domain"	FT	FT	14962..15756	/**tag= ab
FT	misc_RNA	3621..4661	FT	FT	/note= "encodes ketoreductase (KR) 2, the KR domain of module 2"	
FT		/**tag= e	FT	FT	15763..16008	/**tag= ac
FT		/note= "encodes acyl transferase (AT) of the loading domain"	FT	FT	/note= "encodes ACP2, the ACP domain of module 2"	
FT	misc_RNA	4917..5810	FT	FT	16134..37907	/**tag= ad
FT		/**tag= f	FT	FT	/label= epod_gene	
FT		/note= "encodes enoyl reductase (ER) of the loading domain, potentially involved in formation of the thiazole moiety"	FT	FT	/note= "encodes modules 3-6"	
FT	misc_RNA	5856..6155	FT	FT	16425..17606	/**tag= ae
FT		/**tag= g	FT	FT	/note= "encodes KS3"	
FT		/note= "encodes acyl carrier protein (ACP) of the loading domain"	FT	FT	17817..18857	/**tag= af
FT	CDS	6260..10493	FT	FT	/note= "encodes AT3"	
FT		/**tag= h	FT	FT	19581..20396	/**tag= ag
FT		/label= epob_gene	FT	FT	/note= "encodes KR3"	
FT		/note= "encodes module 1, the NRPS module"	FT	FT	20424..20642	/**tag= ah
FT	misc_RNA	2031..3548	FT	FT	/note= "encodes ACP3"	
FT		/note= "encodes condensation domain C2 of the NRPS module"	FT	FT	20706..22082	/**tag= ai
FT		/**tag= i	FT	FT	/note= "encodes KS4"	
FT		/note= "encodes condensation domain C2 of the NRPS module"	FT	FT	22296..23336	/**tag= aj
FT	misc_RNA	2031..3548	FT	FT	/note= "encodes AT4"	
FT		/**tag= j	FT	FT	24069..24647	/**tag= ak
FT		/note= "encodes condensation domain C2 of the NRPS module"	FT	FT	/note= "encodes KR4"	
FT	misc_RNA	6861..6887	FT	FT	24867..25151	/**tag= al
FT		/**tag= k	FT	FT	/note= "encodes ACP4"	
FT		/note= "encodes heterocyclization signature sequence"	FT	FT	25203..26576	/**tag= am
FT	misc_RNA	6861..6887	FT	FT	/note= "encodes KS5"	
FT		/**tag= l	FT	FT	26793..27883	/**tag= an
FT		/note= "encodes condensation domain C4 of the NRPS module"	FT	FT	/note= "encodes AT5"	
FT	misc_RNA	7358..7366	FT	FT	27966..28574	/**tag= ao
FT		/**tag= m	FT	FT	/note= "encodes DH5"	
FT		/note= "encodes condensation domain C7 (partial) of the NRPS module"	FT	FT		
FT	misc_RNA	7898..7921	FT	FT		
FT		/**tag= n	FT	FT		
FT		/note= "encodes adenylation domain A1 of the NRPS module"	FT	FT		
FT	misc_RNA	7898..7921	FT	FT		
FT		/**tag= o	FT	FT		
FT		/note= "encodes adenylation domain A1 of the NRPS module"	FT	FT		
FT	misc_RNA	8261..8308	FT	FT		
FT		/**tag= p	FT	FT		
FT		/note= "encodes adenylation domain A3 of the NRPS module"	FT	FT		
FT	misc_RNA	8411..8422	FT	FT		
FT		/**tag= q	FT	FT		
FT		/note= "encodes adenylation domain A4 of the NRPS module"	FT	FT		
FT	misc_RNA	8861..8905	FT	FT		
FT		/**tag= r	FT	FT		
FT		/note= "encodes adenylation domain A6 of the NRPS module"	FT	FT		
FT	misc_RNA	8966..8983	FT	FT		
FT		/**tag= s	FT	FT		

```
FT misc_RNA 29433..30287
FT /tag= ap
FT /note= "encodes ER5"
FT misc_RNA 30321..30869
FT /tag= aq
FT /note= "encodes KR5"
FT misc_RNA 31077..31373
FT /tag= ar
FT /note= "encodes ACP5"
FT misc_RNA 31440..32807
FT /tag= as
FT /note= "encodes K56"
FT misc_RNA 33018..34067
FT /tag= at
FT /note= "encodes A16"
FT misc_RNA 34107..34676
FT /tag= au
FT /note= "encodes DH6"
FT misc_RNA 35760..36641
FT /tag= av
FT /note= "encodes ER6"
FT misc_RNA 36705..37256
FT /tag= aw
FT /note= "encodes KR6"
FT misc_RNA 37470..37769
FT /tag= ax
FT /note= "encodes ACP6"
FT CDS 37912..49308
FT /tag= ay
FT /label= eproGene
FT /note= "encodes modules 7 and 8"
FT misc_RNA 38014..39375
FT /tag= az
FT /note= "encodes K57"
FT misc_RNA 39589..40626
FT /tag= ba
FT /note= "encodes A17"
FT misc_RNA 41341..41922
FT /tag= db
FT /note= "encodes KR7"
FT misc_RNA 42181..42423
FT

Query Match 100.0%; Score 101; DB 21; Length 71989;
Best Local Similarity 100.0%; Pred. No. 7.5e-18;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGGGATGCGCGCGCGCTGTCATCGCGCGCGGTATCCTCGACGACGCGGCTGCTGA 60
Db 36940 CGGGATGCGCGCGCGCTGTCATCGCGCGGTATCCTCGACGACGCGGCTGCTGA 36999
QY 61 TGCAGCAACCCCGCGCGGTTCCGCGCGTCATGCGGCC 101
Db 37000 TGCAGCAACCCCGCGCGGTTCCGCGCGTCATGCGGCC 37040

RESULT 4
AA255887
ID AA255887 standard; DNA; 68750 BP.
XX AA255887;
XX
XX 10-APR-2000 (first entry)
XX
DE Sorangium cellulosum 68.75 kb contig.
XX
KM Epochlone biosynthesis: type I polyketide synthase; taxol substitute;
KM anticancer; ds.
OS Sorangium cellulosum.
XX Key Location/Qualifiers
FH CDS 1..1826
FT /tag= a
```

```
FT /partial
FT /product= "Partial Orf 1 protein (AAV58580)"
FT /note= "No initiation codon given in the specification"
FT CDS complement (1900..3171)
FT /tag= b
FT /product= "Orf 2 protein (AAV58581)"
FT 3415..5556
FT /tag= c
FT /product= "Orf 3 protein (AAV58582)"
FT CDS complement (5612..5992)
FT /tag= d
FT /product= "Orf 4 protein (AAV58583)"
FT 6226..6675
FT /tag= e
FT /product= "Orf 5 protein (AAV58584)"
FT 7610..11875
FT /tag= f
FT /product= "Type I polyketide synthase, EPOS A (AAV58573)"
FT 11872..116104
FT /tag= g
FT /product= "Non-ribosomal peptide synthetase, EPOS P (AAV58574)"
FT 16251..21749
FT /tag= h
FT /product= "Type I polyketide synthase, EPOS B (AAV58575)"
FT 21746..43519
FT /tag= i
FT /product= "Type I polyketide synthase, EPOS C (AAV58576)"
FT 43524..54920
FT /tag= j
FT /product= "Type I polyketide synthase, EPOS D (AAV58577)"
FT 54935..62254
FT /tag= k
FT /product= "Type I polyketide synthase, EPOS E (AAV58578)"
FT 62369..63628
FT /tag= l
FT /product= "Cytochrome P450 oxygenase homologue, EPOS F (AAV58579)"
FT 63779..64333
FT /tag= m
FT /product= "Orf 6 protein (AAV58585)"
FT CDS complement (63853..64290)
FT /tag= n
FT /product= "Orf 7 protein (AAV58586)"
FT 64363..64920
FT /tag= o
FT /product= "Orf 8 protein (AAV58587)"
FT CDS complement (64287..64727)
FT /tag= p
FT /product= "Orf 9 protein (AAV58588)"
FT 65063..65767
FT /tag= q
FT /product= "Orf 10 protein"
FT CDS complement (65008..65874)
FT /tag= r
FT /product= "Orf 11 protein (AAV58590)"
FT CDS complement (65871..66338)
FT /tag= s
FT /product= "Orf 12 protein (AAV58591)"
FT 66667..67137
FT /tag= t
FT /product= "Orf 13 protein (AAV58592)"
FT CDS 67334..68251
FT /tag= u
FT /product= "Orf 14 protein (AAV58593)"
FT 68346..68750
FT /partial
FT /tag= v
```

```
FT /product= "Partial Orf 15 protein (AAV58594)"
FT /note= "No termination codon given in the specification"
XX
XX
XX MO9966028-A2.
XX
XX 23-DEC-1999.
XX
XX 16-JUN-1999; 99WO-EP04171.
XX
XX 18-JUN-1998; 98US-0099504.
XX 24-SEP-1998; 98US-0101631.
XX 05-FEB-1999; 99US-0118906.
XX
XX (NOVS ) NOVARTIS AG.
XX (NOVS ) NOVARTIS-ERFINDUNGEN VERN GES MBH.
XX
XX Schupp T, Ligon JM, Molnar I, Zirkle R, Goerlach J, Cyr D;
XX
XX WPI: 2000-097741/08.
XX P-PSDB: AAV58573, AAV58574, AAV58575, AAV58576, AAV58577, AAV58578.
XX AAV58579, AAV58580, AAV58581, AAV58582, AAV58583, AAV58584.
XX AAV58585, AAV58586, AAV58587, AAV58588, AAV58590, AAV58591,
XX AAV58592, AAV58593, AAV58594.
XX
XX New isolated epoethlone synthase genes, used for the recombinant
XX production of epoethlone for use in cancer therapy -
XX
XX Claim 14; Page 87-104; 174pp; English.
XX
XX This sequence represents a 68.75 kb contig from Sorangium cellulosum
XX comprising 22 open reading frames (ORFs) and includes genes encoding
XX proteins involved in the biosynthesis of epoethlones. Epoethlones A and
XX B are 16-membered macrocyclic polyketides with an acylcysteine-derived
XX starter unit; polyketides being synthesised from two-carbon building
XX blocks, the beta-carbon of which always carries a keto group. Each round
XX of two-carbon addition is carried out by a complex of enzymes known as
XX the polyketide synthase in a manner similar to fatty acid biosynthesis.
XX EPOS A (AAV58573) and EPOS P (AAV58574) are involved in formation of
XX the thiazole ring formation of epoethlones, and EPOS B, EPOS C, EPOS D
XX and EPOS E (AAV58575-YS8578) are involved in polyketide backbone
XX formation. EPO F (AAV58579) is an epoethlone macroactone oxidase, and
XX the proteins Orf 3 (AAV58582) and Orf14 (AAV58593) are thought to be
XX involved in transport. Epoethlones mimic the biological activity of
XX taxol, and may be substituted for taxol in cancer chemotherapeutic
XX compositions. Epoethlones exhibit a much lower drop in potency against a
XX multiply drug-resistant cell line compared with taxol, and are
XX considerably less efficiently exported from such cells by the multidrug
XX resistance protein (MDR, or P-glycoprotein). Despite the potential of
XX epoethlones as anticancer agents, they are problematical to produce on a
XX large scale. Epoethlones are too complex for industrial scale chemical
XX synthesis, and Sorangium cellulosum is difficult to ferment, producing
XX poor yields of epoethlones. The nucleic acids of the invention may be
XX used for the recombinant production of epoethlones in a heterologous host
XX that is more amenable to fermentation.
XX
XX Sequence 68750 BP; 9596 A; 22458 C; 25537 G; 11159 T; 0 other;
XX
XX Query Match 98.4%; Score 99.4; DB 21; Length 68750;
XX Best Local Similarity 99.0%; Pred. No. 2e-17;
XX Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX 1 CGGGGATCCGCTCGGCGGCTTCATGCGCGGATCGTGGACGACGGGCTGTGGA 60
XX |||||
XX Db 42552 CGGGGATCCGCTCGGCGGCTTCATGCGCGGATCGTGGACGACGGGCTGTGGA 42611
XX
XX 61 TGCAGCAAAACCCCGCGGCTTCGCGCGGATCGATGCGGCC 101
XX |||||
XX Db 42612 TGCAGCAAAACCCCGCGGCTTCGCGCGGATCGATGCGGCC 42652
XX
XX RESULT 5
XX AAF90035
XX ID AAF90035 standard; DNA; 8301 BP.
```

```
XX
XX AAF90035;
XX
XX 06-AUG-2001 (first entry)
XX
XX Nucleotide sequence of a type I polyketide synthase.
XX
XX Metabolic pathway operon: polyketide; polyketide antibiotic;
XX type I polyketide synthase; ss.
XX
XX Unidentified.
XX
XX Key Location/Qualifiers
XX CDS 1..8301
XX FT /tag= a
XX FT /product= "type I polyketide synthase"
XX
XX WO200140497-A2.
XX
XX 07-JUN-2001.
XX
XX 27-NOV-2000; 2000WO-FR03311.
XX
XX 29-NOV-1999; 99FR-0015032.
XX 07-JUN-2000; 2000US-0209800.
XX
XX (AVET ) AVENTIS PHARMA SA.
XX
XX Jeanin P, Pernodet J, Guerin M, Simonet P, Courtois S;
XX Cappellano C, Francon F, Raynal A, Ball M, Sezonov G, Tuphile K;
XX Frostegard A;
XX
XX WPI: 2001-374849/39.
XX P-PSDB: AAB83972.
XX
XX Collection of nucleic acids from environmental samples, useful for
XX identifying e.g. genes encoding polyketide synthases and derived
XX antibiotics -
XX
XX Claim 35; Page 302-305; 356pp; French.
XX
XX The specification describes a method for the preparation of a collection
XX of nucleic acids from organisms in a soil sample. The method comprises
XX milling a dried sample to produce microparticles; suspending these in
XX liquid buffer; extraction of nucleic acids from the microparticle;
XX passing nucleic acid-containing solution through an anion exchange
XX chromatography material; and recovering fractions containing purified
XX nucleic acids. The nucleic acids are sources for sequences that encode
XX either operons involved in a metabolic pathway (specifically polyketide
XX synthesis) or polypeptides, particularly for production of therapeutic
XX or agricultural compounds, especially polyketide antibiotics. AAF90034-39
XX represent open reading frames (ORFs) of the coding strand of cosmid
XX a26g1, and encode type I polyketide synthases.
XX
XX Sequence 8301 BP; 1426 A; 2629 C; 2669 G; 1577 T; 0 other;
XX
XX Query Match 46.7%; Score 47.2; DB 22; Length 8301;
XX Best Local Similarity 69.6%; Pred. No. 0.0012;
XX Matches 64; Conservative 0; Mismatches 28; Indels 0; Gaps 0;
XX
XX 9 CGCGTCCGCGGCGGTCATGCGCGGATCGTATCCGAGACGACGGGCTGTGATGACGAA 68
XX |||||
XX Db 7435 CGCGTCCGCGGCGGTCATGCGCGGATCGTATCCGAGACGACGGGCTGTGATGACG 7494
XX
XX 69 ACCCCGCGCGGATTCGCGCGGATCGATGCGGCC 100
XX |||||
XX Db 7495 GACGTGACGCGGATCGCAAGGCTATGCGGCC 7526
XX
XX RESULT 6
XX AAF90033
XX ID AAF90033 standard; DNA; 34071 BP.
```

```
XX AAF90033;
AC
XX
OS
XX 06-AUG-2001 (first entry)
DT
XX
DE Nucleotide sequence of cosmid a26g1 (coding strand).
XX
XX Metabolic pathway operon; polyketide; polyketide antibiotic; ss.
XX
XX Synthetic.
OS
XX WO200140497-A2.
PN
XX
XX 07-JUN-2001.
PD
XX
XX 27-NOV-2000; 2000MO-FR03311.
PF
XX
XX 29-NOV-1999; 99FR-0015032.
PR
XX 07-JUN-2000; 2000US-0209800.
XX
XX (AVET ) AVENTIS PHARMA SA.
PA
XX
PI Jeanin P, Pernodet J, Guerneau M, Simonet P, Courtois S;
PI Cappellano C, Francou F, Raynal A, Ball M, Sezonov G, Tuphille K;
PI Frostegard A;
XX
XX WPI; 2001-374849/39.
DR
XX
XX Collection of nucleic acids from environmental samples, useful for
PT identifying e.g. genes encoding polyketide synthases and derived
PT antibiotics .
XX
XX Example 14; Page 289-300; 356pp; French.
PS
XX
XX The specification describes a method for the preparation of a collection
CC of nucleic acids from organisms in a soil sample. The method comprises
CC milling a dried sample to produce microparticles; suspending these in
CC liquid buffer; extraction of nucleic acids from the microparticle;
CC passing nucleic acid-containing solution through a molecular sieve;
CC passing nucleic acid-enriched fractions through an anion exchange
CC chromatography material; and recovering fractions containing purified
CC nucleic acids. The nucleic acids are sources for sequences that encode
CC either operons involved in a metabolic pathway (specifically polyketide
CC synthetase) or polypeptides, particularly for production of therapeutic
CC or agricultural compounds, especially polyketide antibiotics. The present
CC sequence represents cosmid a26g1 coding strand), which encodes different
CC polyketide synthases.
XX
XX Sequence 34071 BP; 5791 A; 10858 C; 11089 G; 6333 T; 0 other;
SQ
Query Match 46.7%; Score 47.2; DB 22; Length 34071;
Best Local Similarity 69.6%; Pred. No. 0.0012;
Matches 64; Conservative 0; Mismatches 28; Indels 0; Gaps 0;
QY 9 CCGCTCCGCGGCGCTTCATGCGCGCGGTATCTGGACGACGGCGTGTGATGACGCAA 68
DB 12067 CCGCTCGCGGCGCTTCATGCGCGCGGTGTCTGGACGACGGCGTGTGATGACG 12126
QY 69 ACCCCGCGGCGGTTCGCGCGGCGATGCGCGC 100
DB 12127 GACTGGACGCGGATCCAGACGTCATGCGCGC 12158
RESULT 7
AAAF90032/c
ID AAF90032 standard; DNA; 42717 BP.
XX
XX AAF90032;
AC
XX
XX 06-AUG-2001 (first entry)
DT
XX
XX Nucleotide sequence of cosmid a26g1 (non-coding strand).
DE
XX
```

```
KW Metabolic pathway operon; polyketide; polyketide antibiotic; ss.
XX
XX Synthetic.
OS
XX WO200140497-A2.
PN
XX
XX 07-JUN-2001.
PD
XX
XX 27-NOV-2000; 2000MO-FR03311.
PF
XX
XX 29-NOV-1999; 99FR-0015032.
PR
XX 07-JUN-2000; 2000US-0209800.
XX
XX (AVET ) AVENTIS PHARMA SA.
PA
XX
XX Jeanin P, Pernodet J, Guerneau M, Simonet P, Courtois S;
PI Cappellano C, Francou F, Raynal A, Ball M, Sezonov G, Tuphille K;
PI Frostegard A;
XX
XX WPI; 2001-374849/39.
DR
XX
XX Collection of nucleic acids from environmental samples, useful for
PT identifying e.g. genes encoding polyketide synthases and derived
PT antibiotics .
XX
XX Example 14; Page 274-288; 356pp; French.
PS
XX
XX The specification describes a method for the preparation of a collection
CC of nucleic acids from organisms in a soil sample. The method comprises
CC milling a dried sample to produce microparticles; suspending these in
CC liquid buffer; extraction of nucleic acids from the microparticle;
CC passing nucleic acid-containing solution through a molecular sieve;
CC passing nucleic acid-enriched fractions through an anion exchange
CC chromatography material; and recovering fractions containing purified
CC nucleic acids. The nucleic acids are sources for sequences that encode
CC either operons involved in a metabolic pathway (specifically polyketide
CC synthetase) or polypeptides, particularly for production of therapeutic
CC or agricultural compounds, especially polyketide antibiotics. The present
CC sequence represents cosmid a26g1 (non-coding strand). The sense strand
CC encodes different polyketide synthases.
XX
XX Sequence 42717 BP; 8230 A; 13520 C; 13184 G; 7782 T; 1 other;
SQ
Query Match 46.7%; Score 47.2; DB 22; Length 42717;
Best Local Similarity 69.6%; Pred. No. 0.0012;
Matches 64; Conservative 0; Mismatches 28; Indels 0; Gaps 0;
QY 9 CCGCTCCGCGGCGCTTCATGCGCGCGGTATCTGGACGACGGCGTGTGATGACGCAA 68
DB 22053 CCGCTCGCGGCGCTTCATGCGCGCGGTGTCTGGACGACGGCGTGTGATGACG 21994
QY 69 ACCCCGCGGCGGTTCGCGCGGCGATGCGCGC 100
DB 21993 GACTGGACGCGGATCCAGACGTCATGCGCGC 21962
RESULT 8
AAAF88339
ID AAF88339 standard; DNA; 16767 BP.
XX
XX AAF88339;
AC
XX
XX 28-AUG-2001 (first entry)
DT
XX
XX S. spinosa DNA fragment encoding ORF22, SEQ ID 49.
DE
XX
XX Forosamine; trimethylrhamnose; polyketide synthase; biosynthesis;
KW spinosyn; polyketide aglycone; transgenic plant; insect resistance;
KW macrolide; insecticidal; polyketide synthase; ds.
XX
XX Saccharopolyspora spinosa.
OS
XX
XX DE19957268-A1.
PN
```

[illegible][illegible]

PA (FARB) BAYER AG.
 XX Eberz G, Moehle V, Froede R, Velten R, Salas JA:
 XX WPI, 2001-267102/28.
 DR
 XX
 PT New nucleic acid encoding enzymes for spinosyn biosynthesis, useful for
 PT recombinant production of insecticidal spinosyns and their derivatives
 PT
 PS
 Claim 7: Page 31-49; 354pp; German.
 XX
 CC This invention describes a novel method nucleic acid (I) and its encoded
 CC polypeptide (II) containing at least one region that encodes an enzymatic
 CC activity involved in biosynthesis of spinosyns. (I) are used (1) to
 CC identify, inactive or modulate genes involved in the biosynthesis of
 CC (II); (11) to generate a library of polypeptide synthetases; (111) for
 CC adding forosamine or trimethylrhinose to a spinosyn or polypeptide
 CC aglycone; and (IV) for recombinant production of the corresponding
 CC derivatives, including production of transgenic plants that express (II)
 CC and thus have increased resistance to insects. (I) are also useful as
 CC markers for sequencing of the Saccharopolyspora spinosa genome. (II) are
 CC also be used to raise specific antibodies, useful for identifying
 CC expression clones in a gene bank. Cells transformed with (I) may produce
 CC (II) at significantly increased levels or produce new derivatives of
 CC (II). This sequence represents a genomic DNA fragment of the S. spinosa
 CC genome which contains the coding regions for proteins involved in
 CC forosamine, trimethylrhinose and polypeptide synthase biosynthesis.
 CC
 XX
 SQ Sequence 50000 BP; 6867 A; 14165 C; 19274 G; 9694 T; 0 other;
 Query Match 44.6%; Score 45; DB 22; Length 50000;
 Best Local Similarity 65.3%; Pred. No. 0.0048;
 Matches 66; Conservative 0; Mismatches 35; Indels 0; Gaps 0;
 QY 1 CGGGGATGCCGTCGCGGCGTCGTCATGCGCGGTATCTGAGACGAGCGGCTGCTGA 60
 DB 49460 CGGAAACCCGTCGCGGTGTCGTCACGCCGCGGTGTCGATGACGTCGCTGA 49519
 QY 61 TGCAGCAACCCCGCGGTCGCCGCGTCATGCGGCC 101
 DB 49520 TGTGATGTCGCCGAGCGCTTGACGCGGTCGCGGCC 49560
 RESULT 11
 AA221501
 ID AA221501 standard; DNA: 80161 BP.
 AC AA221501;
 XX
 DT 01-DEC-1999 (first entry)
 XX
 DE DNA fragment of Saccharopolyspora spinosa containing biosynthetic genes.
 XX
 KW spinosyn biosynthetic enzyme; open reading frame; ORF; insecticidal
 KW microtides; arachnid; nematode; insect; polypeptide; polypeptide synthase;
 KW PKS; extender module; initiator module; acyl transferase domain; AT;
 KW acyl carrier protein; AC; beta-ketosynthase domain; KS; KR;
 KW dehydratase domain; DH; enoyl reductase domain; ER; beta-ketoreductase;
 KW insecticide; ss.
 XX
 OS Saccharopolyspora spinosa.
 XX
 FH Key Location/Qualifiers
 FT CDS complement (1135..1971)
 FT /tag- a
 FT /product= ORFL16
 FT /note= "Protein involved in transcription control"
 FT 2024..2791
 FT /tag- b
 FT /product= ORFL15

FT /note= "keto acyl reductase"
 FT complement (3416..4165)
 FT /tag- c
 FT /product= spns
 FT /note= "spinosyn biosynthesis protein S"
 FT complement (4168..5325)
 FT /tag- d
 FT /product= spnr
 FT /note= "spinosyn biosynthesis protein R"
 FT complement (5363..6751)
 FT /tag- e
 FT /product= spnq
 FT /note= "spinosyn biosynthesis protein Q"
 FT 7083..8450
 FT /tag- f
 FT /product= spnp
 FT /note= "spinosyn biosynthesis protein P"
 FT 8967..10427
 FT /tag- g
 FT /product= spno
 FT /note= "spinosyn biosynthesis protein O"
 FT 10436..11434
 FT /tag- h
 FT /product= spnn
 FT /note= "spinosyn biosynthesis protein N"
 FT complement (11530..12492)
 FT /tag- i
 FT /product= spnm
 FT /note= "spinosyn biosynthesis protein M"
 FT complement (12696..13547)
 FT /tag- j
 FT /product= spnl
 FT /note= "spinosyn biosynthesis protein L"
 FT complement (13592..14785)
 FT /tag- k
 FT /product= spnk
 FT /note= "spinosyn biosynthesis protein K"
 FT complement (14799..16418)
 FT /tag- l
 FT /product= spnj
 FT /note= "spinosyn biosynthesis protein J"
 FT 16556..17743
 FT /tag- m
 FT /product= spni
 FT /note= "spinosyn biosynthesis protein I"
 FT complement (17749..18501)
 FT /tag- n
 FT /product= spnh
 FT /note= "spinosyn biosynthesis protein H"
 FT complement (18541..19713)
 FT /tag- o
 FT /product= spng
 FT /note= "spinosyn biosynthesis protein G"
 FT 20168..20995
 FT /tag- p
 FT /product= spnf
 FT /note= "spinosyn biosynthesis protein F"
 FT 21111..28898
 FT /tag- q
 FT /product= spna
 FT /note= "spinosyn biosynthesis protein A"
 FT /transl_except= (Pos:26940..26942, aa:Pro)
 FT 21126..22379
 FT /tag- r
 FT /note= "Beta-ketosynthase domain (KS1): part of the
 FT initiator module"
 FT 22692..23669
 FT /tag- s
 FT /note= "Acyl transferase domain (AT1): part of the
 FT initiator module"
 FT 23793..24041
 FT /tag- t
 FT /note= "Acyl carrier protein domain (ACPI): part of the
 FT msc_feature

```
FT misc-feature 24102..25349 initiator module"
FT /tag= u
FT /note= "Beta-ketosynthase domain (KS1): part of
FT extender module 1"
FT misc-feature 25683..26684
FT /tag= v
FT /note= "Acyl transferase domain (AT1): part of
FT extender module 1"
FT misc-feature 27582..28121
FT /tag= w
FT /note= "Beta-ketoreductase domain (KR1): part of
FT extender module 1"
FT misc-feature 28404..28649
FT /tag= x
FT /note= "Acyl carrier protein domain (ACP1): part of
FT extender module 1"
FT CDS 38916..35374
FT /tag= y
FT /product= spnc
FT /note= "Spinosyn biosynthesis protein C"
FT misc-feature 29024..30295
FT /tag= z
FT /note= "Beta-ketosynthase domain (KS2): part of
FT extender module 2"
FT misc-feature 30629..31621
FT /tag= aa
FT /note= "Acyl transferase domain (AT2): part of
FT extender module 2"
FT misc-feature 31697..32254
FT /tag= ab
FT /note= "Dehydratase domain (DH2): part of extender
FT module 2"
FT misc-feature 33035..34072
FT /tag= ac
FT /note= "Enoyl reductase domain (ER2): part of
FT extender module 2"
FT misc-feature 34082..34621
FT /tag= ad
FT /note= "Beta-ketoreductase domain (KR2): part of
FT extender module 2"
FT misc-feature 34886..30295
FT /tag= ae
FT /note= "Acyl carrier protein domain (ACP2): part of
FT extender module 2"
FT CDS 35419..44931
FT /tag= af
FT /product= spnc
FT /note= "Spinosyn biosynthesis protein C"
FT misc-feature 35518..36786
FT /tag= ag
FT /note= "Beta-ketosynthase domain (KS3): part of
FT extender module 3"
FT misc-feature 37108..38097
FT /tag= ah
FT /note= "Acyl transferase domain (AT3): part of
FT extender module 3"
FT misc-feature 38992..39528
FT /tag= ai
FT /note= "Beta-ketoreductase domain (KR3): part of
FT extender module 3"
FT misc-feature 39790..40035
FT /tag= aj
FT /note= "Acyl carrier protein domain (ACP3): part of
FT extender module 3"
FT misc-feature 40102..41373
FT /tag= ak
FT /note= "Beta-ketosynthase domain (KS4): part of
FT extender module 4"
FT misc-feature 41713..42705
FT /tag= al
FT /note= "Acyl transferase domain (AT4): part of
FT extender module 4"
```

```
FT misc-feature 43615..44157
FT /tag= am
FT /note= "Beta-ketoreductase domain (KR4): part of
FT extender module 4"
FT misc-feature 44431..44676
FT /tag= an
FT /note= "Acyl carrier protein domain (ACP4): part of
FT extender module 4"
FT CDS 44966..59752
FT /tag= ao
FT /product= spnd
FT /note= "Spinosyn biosynthesis protein D"
FT misc-feature 45077..46348
FT /tag= ap
FT /note= "Beta-ketosynthase domain (KS5): part of
FT extender module 5"
FT misc-feature 46691..47674
FT /tag= aq
FT /note= "Acyl transferase domain (AT5): part of
FT extender module 5"
FT misc-feature 47753..48310
FT /tag= ar
FT /note= "Dehydratase domain (DH5): part of extender
FT module 5"
FT misc-feature 49226..49771
FT /tag= as
FT /note= "Beta-ketoreductase domain (KR5): part of
FT extender module 5"
FT misc-feature 50009..50254
```

Query Match 44.6%; Score 45; DB 20; Length 80161;
Best Local Similarity 65.3%; Pred. No. 0.0049;
Matches 66; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

```
QY 1 CGGGGATGCCGCTCCGCGCGCTTCATGCGCCGCGATTCGTGACGACGCGGCTGCTGA 60
DB 74831 CGGAACCCGTTGGGGGTGGTGACGCGCGCGTGTCTGTGATGACGGTGTGCTGA 74890
QY 61 TGCAGCAACCCCGCGCGGTTCCGCGCGGTATGCGGCC 101
DB 74891 TGTGATGTGCGCGGACGCGCTTGACGCGGTGTGCGGCC 74931
```

```
RESULT 12
AAA54013
ID AAA54013 standard; DNA; 7131 BP.
AC AAA54013;
XX
AC 08-FEB-2001 (first entry)
XX
DE 6-Methylsalicylic acid coding sequence.
XX
KW Disease; fungi; bacteria; virus; nematode; plant pathogen;
KW polyketide synthase; PK; 6-methylsalicylic acid; phosphopantetheinyl;
KW crop protection; resistance; systemic acquired resistance; SAR;
KW maize; wheat; rice; barley; sorghum; oats; rye; soybean; Brassica sp;
KW sunflower; safflower; alfalfa; potato; peanut; cotton; ds.
XX
OS Penicillium patulum.
XX
PN WO20005340-A1.
XX
PD 21-SEP-2000.
XX
PF 24-FEB-2000; 2000WO-US04691.
XX
PR 15-MAR-1999; 99US-0124374.
XX
PA (PION-) PIONEER HI-BRED INT INC.
XX
PI Yajpant N;
```

DR WPI; 2000-638204/61.
 XX Enhancing resistance of plants against pathogens e.g. fungi, bacteria,
 PT viruses and nematodes, involves incorporating a DNA construct encoding
 PT polyketide synthase into the plant genome to increase defense-related
 PT protein levels
 XX Claim 6; Page 50-52; 68pp; English.
 XX
 CC Disease in plants is caused by biotic and abiotic causes. Biotic
 CC causes include fungi, viruses, bacteria and nematodes and of these
 CC fungi are the most causative agent of disease on plants. A host of
 CC cellular processes enable plants to defend themselves from disease
 CC caused by pathogenic agents. In addition to the localised
 CC hypersensitive response, plants have evolved a systemic defense
 CC mechanism that reduces the impact of subsequent pathogen attacks.
 CC The accumulation of a group of extracellular proteins called
 CC pathogenesis related (PR) proteins were reported to correlate with
 CC the onset of the systemic acquired resistance. Salicylic acid, a
 CC plant produced phenolic compound was implicated as a signal in
 CC systemic acquired resistance. Thus, new methods of enhancing disease
 CC resistance in a plant involves stably incorporating a DNA construct,
 CC comprising a sequence encoding a polyketide synthase (PS) such as
 CC 6-methylsalicylic acid, to a promoter capable of driving gene
 CC expression, into the plant genome, so that the level of
 CC defense-related protein is increased. Alternatively, stably
 CC transforming a plant cell with a polyketide synthase gene and an
 CC additional gene which encodes a protein that, for example, would
 CC increase the level of substrate for the polyketide synthase or
 CC convert the polyketide synthase from an inactive to an active form.
 CC The method is useful for enhancing disease resistance in plants and
 CC commercially viable crops such as maize, wheat, barley, sorghum,
 CC oats, rye, soybean, Brassica sp., sunflower, safflower, alfalfa,
 CC potato, peanut and cotton. The method enhances resistance against
 CC pathogens including fungi, bacteria, viruses and nematodes, and has
 CC applications in agriculture. Advantageously, polyketide synthase
 CC genes can be used to transform any desired plant.
 XX
 SQ Sequence 7131 BP; 1610 A; 2024 C; 1865 G; 1632 T; 0 other;
 Query Match 43.6%; Score 44; DB 21; Length 7131;
 Best Local Similarity 67.4%; Pred. No. 0.0082;
 Matches 62; Conservative 0; Mismatches 30; Indels 0; Gaps 0;
 QY 10 CGCTCCGCGCGCTGCTATGCGCGGTATCCTGACGACGCTGATGACGAA 69
 DB 5415 CGGTCCAGGGTGTGTTACAGCTGCGGTGCTGCGACACGACTAGTATGACGCA 5474
 QY 70 CCCCCGCGCGCTTCCGCGCGGTATGCGCGCC 101
 DB 5475 CTCGCGACGCTTCAACCGCGGTCTCGACCC 5506
 RESULT 13
 AAD17186
 ID AAD17186 standard; DNA; 125401 BP.
 XX
 AC AAD17186;
 XX
 DT 29-NOV-2001 (first entry)
 XX
 DE Streptomyces noursei nystatin PKS gene cluster DNA.
 XX
 KW Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
 KM antifungal; antibiotic; ds.
 XX
 OS Streptomyces noursei.
 XX
 FH Key Location/Qualifiers
 FT CDS 6337..34771
 FT /*tag= "NysI complete protein"
 FT /product= "NysI complete protein"
 FT CDS 34792..51099

FT /*tag= b
 FT /product= "NysJ protein"
 FT CDS 51155..57355
 FT /*tag= c
 FT /product= "NysK protein"
 FT CDS 57503..58687
 FT /*tag= d
 FT /product= "NysL protein"
 FT CDS complement (58786..58980)
 FT /*tag= e
 FT /product= "NysM protein"
 FT note= "CDS does not include start codon"
 FT complement (59045..60241)
 FT /*tag= f
 FT /product= "NysN protein"
 FT note= "CDS does not include start codon"
 FT complement (60238..61296)
 FT /*tag= g
 FT /product= "NysD2 complete protein"
 FT CDS 120628..121308
 FT /*tag= h
 FT /product= "NysR4 (long) protein"
 XX
 PN W0200159126-A2.
 XX
 PD 16-AUG-2001.
 XX
 XX 08-FEB-2001; 2001MWO-GB00509.
 XX
 PF 08-FEB-2000; 2000GB-0002840.
 XX
 PR 10-APR-2000; 2000GB-0008786.
 PR 14-APR-2000; 2000GB-0009387.
 XX
 PA (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
 PA (SNPE) SINTEF STREITELSEN IND TEK FORSK.
 PA (ALPR-) ALPHARMA AS.
 PA (SINV-) SINVENT AS.
 PA (DZIE/) DZIEGLEWSKA H.
 PA (ZOTC/) ZOTCHEV S B.
 PA (SEKU/) SEKUROVA O N.
 PA (FJAE/) FJAEVRIK E.
 PA (BRAU/) BRAUTASET T.
 PA (STRO/) STROM A R.
 XX
 PI Zotchev SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AR;
 PI Valla S, Ellingsen TE, Sletta H, Gulliksen O;
 DR WPI; 2001-557614/62.
 DR P-PSDB: AAE10143, AAE10144, AAE10145, AAE10146, AAE10147, AAE10148,
 DR AAE10149, AAE10150.
 PT New nystatin polyketide synthase polynucleotides and polypeptides,
 PT useful as antibiotics and antifungals -
 XX
 XX Claim 1; Page 188-254; 266pp; English.
 PS
 XX The present invention relates to the cloning and sequencing of the gene
 CC cluster encoding a modular type I polyketide synthase (PKS) enzyme
 CC involved in the biosynthesis of the macrolide antibiotic nystatin.
 CC The nystatin PKS is useful as antifungal antibiotics. The present
 CC sequence is a Streptomyces noursei nystatin PKS gene cluster DNA.
 XX
 SQ Sequence 125401 BP; 15664 A; 49692 C; 42871 G; 17174 T; 0 other;
 Query Match 43.6%; Score 44; DB 22; Length 125401;
 Best Local Similarity 70.2%; Pred. No. 0.0091;
 Matches 59; Conservative 0; Mismatches 25; Indels 0; Gaps 0;
 QY 8 GCGGTCGCGCGGCTGCTATGCGCGGTATCCTGACGACGCGCTGATGACGCA 67
 DB 33721 GCGGTCGACCGGTGTGTGTCACACCGCGGTCTCTGACGACGCGTACTGACCGGCT 33780
 QY 68 AACCCCGCGCGGCTTCCGCGCGGT 91

score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

(without alignments)
6625.829 Million cell updates/sec

Sequence: 1 **cggggatgccgctccg**gc.....**tcgcgcggtcatg**gcgcc 101

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Total number of hits satisfying chosen parameters: 4109280

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

1:	GenEmb1: *
2:	gb_ba: *
3:	gb_hhg: *
4:	gb_in: *
5:	gb_om: *
6:	gb_ov: *
7:	gb_pat: *
8:	gb_ph: *
9:	gb_pl: *
10:	gb_pr: *
11:	gb_ro: *
12:	gb_sts: *
13:	gb_sy: *
14:	gb_un: *
15:	gb_vt: *
16:	em_ba: *
17:	em_fun: *
18:	em_hum: *
19:	em_in: *
20:	em_mu: *
21:	em_com: *
22:	em_or: *
23:	em_ov: *
24:	em_pat: *
25:	em_pl: *
26:	em_ro: *
27:	em_sts: *
28:	em_un: *
29:	em_vt: *
30:	em_hhg_hum: *
31:	em_hhg_in: *
32:	em_hhg_other: *
33:	em_hhg_mus: *
34:	em_hhg_pln: *
35:	em_hhg_pro: *
36:	em_hhg_mam: *
37:	em_hhg_vrt: *
38:	em_sy: *
39:	em_hhg_hum: *
40:	em_hhg_mus: *
41:	em_hhg_other: *

Pred. No. is the number of results predicted by chance to have a

Result No.	Score	Query Match	Length	DB	ID	Description
1	101	100.0	552	6	AX403009	AX403009 Sequence
2	101	100.0	552	6	AX403011	AX403011 Sequence
3	101	100.0	58733	1	AE217189	AE217189 Sorangium
4	101	100.0	71989	6	AE127664	AE127664 Sequence
5	99.4	98.4	68750	1	AE210843	AE210843 Sorangium
6	99.4	98.4	68750	6	AR193028	AR193028 Sequence
7	99.4	98.4	68750	6	AR193551	AR193551 Sequence
8	99.4	98.4	68750	6	AR199559	AR199559 Sequence
9	99.4	98.4	68750	6	AR199567	AR199567 Sequence
10	99.4	98.4	68750	6	AR201097	AR201097 Sequence
11	99.4	98.4	68750	6	AR208671	AR208671 Sequence
12	76	75.2	16124	1	AX024384	AX024384 Sequence
13	76	75.2	16124	6	AX024277	AX024277 Sequence
14	53	52.5	49736	1	AF319998	AF319998 Stigmatel
15	48.2	47.7	3225	1	AF232752	AF232752 Mycobacte
16	48.2	47.7	24225	1	SC264	AL512902 Streptomy
17	47.2	46.7	8301	6	AX153792	AX153792 Sequence
18	47.2	46.7	34071	6	AX153790	AX153790 Sequence
19	47.2	46.7	42603	1	AF188287	AF188287 Stigmatel
20	47.2	46.7	42717	6	AX153789	AX153789 Sequence
21	45.2	44.8	82746	1	AF453501	AF453501 Actinosyr
22	45	44.6	16767	6	AX089464	AX089464 Sequence
23	45	44.6	29736	6	AX089421	AX089421 Sequence
24	45	44.6	50000	6	AX089417	AX089417 Sequence
25	45	44.6	80161	1	AF007564	AF007564 Saccharop
26	45	44.6	80161	6	AR165018	AR165018 Sequence
27	44.4	44.0	104326	1	AB070940	AB070940 Streptomy
28	44	43.6	7131	8	PPMSAS	X55779 Penicillium
29	44	43.6	25150	1	AB070945	AB070945 Streptomy
30	44	43.6	123580	1	AF263912	AF263912 Streptomy
31	44	43.6	125401	6	AX211779	AX211779 Sequence
32	43.8	43.4	20394	1	SNM132222	AJ132222 Streptomy
33	43.8	43.4	20394	6	AX067996	AX067996 Sequence
34	43.8	43.4	84985	1	SMA278573	AJ278573 Streptomy
35	43.4	43.0	66808	1	SAU21825	AJ421825 Stigmatel
36	43	42.6	27522	1	AB070942	AB070942 Streptomy
37	42.4	42.0	30000	6	AX250262	AX250262 Sequence
38	41.8	41.4	5292	6	AX153793	AX153793 Sequence
39	41.8	41.4	6462	6	AX153794	AX153794 Sequence
40	41.8	41.4	12381	6	AX006889	AX006889 Sequence
41	41.8	41.4	31690	6	E38020	Avermectin
42	41.8	41.4	31422	6	E38021	Avermectin
43	41.8	41.4	39314	1	SGR300302	AJ300302 Streptomy
44	41.8	41.4	41097	1	AF016585	AF016585 Streptomy
45	41.8	41.4	64957	1	AB032367	AB032367 Streptomy

RESULT 1
AX403009

LOCUS	552 bp	DNA	linear	PAT 07-JUN-2002
AX403009				
DEFINITION	Sequence 32 from Patent WO0183800.			

VERSION AX403009.1 GI:21387990

SOURCE synthetic construct.

artificial sequences

AUTHOR

TITLE
JOURNAL
Khosla, C., Lau, J., Licardi, P.J., Regenstein, R., Santil, D. and Yang, L.
Heterologous production of polyketides
Patent: WO 0183800-A 32 08-NOV-2001;

FEATURES Kosan Biosciences, Inc. (US)
Location/Qualifiers
source
1..552
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="DNA sequence of the KR domain of extender module 6 of the epothilone PKS"
<1..>552
/note="unnamed protein product"
/codon_start=1
/transl_table=11
/protein_id="CAD33696.1"
/db_xref="GI:21387991"
/translation="DGYLVYTGAGLGLISVAGWLAEOGAGHLVTVGRSGAVSAEQOT
AAALFAGARVTVARADVDRAQIERILREVTASGMPLRGVYVHAAGIDDLMOOT
PARFRVAVAPKVRGALHLHALREAPPLSFVLYASGAGLLSGPGGNVAAATFIDAL
AHHRAOGPLALSDMLGLEADVGL"

BASE COUNT 70 a 176 c 214 g 92 t

ORIGIN

Query Match 100.0%; Score 101; DB 6; Length 552;
Best Local Similarity 100.0%; Pred. No. 7.8e-12;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGGGATGCCGCTCGCGCGCTTCATGCGCGGTATCCTGACGACGGCGTCTGA 60
|||||
Db 236 CGGGATGCCGCTCGCGCGCTTCATGCGCGGTATCCTGACGACGGCGTCTGA 295
|||||

QY 61 TGCAGCAACCCCGCGCGCTTCGCGCGGTATGCGGCC 101
|||||

Db 296 TGCAGCAACCCCGCGCGCTTCGCGCGGTATGCGGCC 336
|||||

RESULT 2
AX403011 552 bp DNA linear PAT 07-JUN-2002
LOCUS Sequence 34 from Patent WO0183800.
DEFINITION AX403011
ACCESSION AX403011
VERSION AX403011.1 GI:21387992
KEYWORDS
SOURCE
ORGANISM
synthetic construct.
synthetic construct.
artificial sequences.
REFERENCE
1
AUTHORS
Arslanian, R.L., Ashley, G., Frykman, S., Julien, B., Katz, L.,
Khosla, C., Lau, J., Licardi, P.J., Regentlin, R., Santl, D. and Tang, L.
TITLE
Heterologous production of polyketides
JOURNAL
Patent: WO 0183800-A 34 08-NOV-2001;
Kosan Biosciences, Inc. (US)
Location/Qualifiers
source
1..552
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="DNA sequence of the mutated KR domain of extender
module 6 of the novel 9-keto-epothilone PKS"
<1..>552
/note="unnamed protein product"
/codon_start=1
/transl_table=11
/protein_id="CAD33697.1"
/db_xref="GI:21387993"
/translation="DGYLVYTGAGLGLISVAGWLAEOGAGHLVTVGRSGAVSAEQOT
AAALFAGARVTVARADVDRAQIERILREVTASGMPLRGVYVHAAGIDDLMOOT
PARFRVAVAPKVRGALHLHALREAPPLSFVLYASGAGLLSGPGGNVAAATFIDAL
AHHRAOGPLALSDMLGLEADVGL"

BASE COUNT 70 a 177 c 212 g 93 t

ORIGIN

Query Match 100.0%; Score 101; DB 6; Length 552;
Best Local Similarity 100.0%; Pred. No. 7.8e-12;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGGGATGCCGCTCGCGCGCTTCATGCGCGGTATCCTGACGACGGCGTCTGA 60
|||||

Db 236 CGGGATGCCGCTCGCGCGCTTCATGCGCGGTATCCTGACGACGGCGTCTGA 295
|||||

QY 61 TGCAGCAACCCCGCGCGCTTCGCGCGGTATGCGGCC 101
|||||

Db 296 TGCAGCAACCCCGCGCGCTTCGCGCGGTATGCGGCC 336
|||||

RESULT 3
AF217189 58733 bp DNA linear BCT 09-JUN-2000
LOCUS Sorangium cellulosum putative transposase gene, partial cds;
DEFINITION putative transposase gene, complete cds; epothilone biosynthesis
gene cluster, complete sequence; putative membrane protein gene,
complete cds.
ACCESSION AF217189
VERSION AF217189.1 GI:7453554
KEYWORDS
Polyangium cellulosum.
ORGANISM
Polyangium cellulosum.
Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
Myxococcales; Sorangineae; Polyangiaceae; Polyangium.
REFERENCE
1 (bases 1 to 58733)
AUTHORS
Tang, L., Shah, S., Chung, L., Carney, J., Katz, L., Khosla, C. and
Julien, B.
TITLE
Cloning and heterologous expression of the epothilone gene cluster
JOURNAL
Science 287 (5453), 640-642 (2000)
MEDLINE
20115955
PUBMED
10649995
REFERENCE
2 (bases 1 to 58733)
AUTHORS
Julien, B., Shah, S., Ziermann, R., Goldman, R., Katz, L. and Khosla, C.
TITLE
Isolation and characterization of the epothilone biosynthetic gene
cluster from Sorangium cellulosum
JOURNAL
Gene 249 (1-2), 153-160 (2000)
MEDLINE
20293058
PUBMED
10831849
REFERENCE
3 (bases 1 to 58733)
AUTHORS
Julien, B.
TITLE
Direct Submission
JOURNAL
Submitted (16-DEC-1999) Kosan Biosciences, Inc., 3832 Bay Center
Place, Hayward, CA 94545, USA
Location/Qualifiers
source
1..58733
/organism="Polyangium cellulosum"
/strain="SMP4"
/db_xref="taxon:56"
<1..992
/note="orfA"
/codon_start=3
/transl_table=11
/product="putative transposase"
/protein_id="AAF62878.1"
/db_xref="GI:7453555"
/translation="VRGHEAADAGGCPRVLLYNLKNAYVERGDAIRRPITLAL
SADYRPPVAVARGNEKGRVERAIRYREGFEFARAADVGLNRQATEWTSAL
DSWVEDRARTVROAFDDERSVLLRHPDPPEDEHREVEVGKPTYPARDLDYVPH
DTRRTVLVADLSQVRIADGNQIVATVHRSRDQQLQEPHDLRLVDEKRAEHR
GDLRLAARARSQAFLRIYAERGDVNGSAIARLLDLDAVGALEALVEALVEEDTI
HTGAVROVDRRRERHLEPPVSIYTRREHALVVTPLSLTYALKKDPF"
989..1501
/note="orfB"
/codon_start=1
/transl_table=11
/product="putative transposase"
/protein_id="AAF62879.1"
/db_xref="GI:7453556"
/translation="MTDLPTETKDLKSLGLFGLLACBEOADKPMULREVLAIEERE
RHKRSIERLKNRYAAAFKPMTPDPSWPKKIDRAVVDLYSRADLLEFVETRYD
AOKPLLSITNKAFAWMGVFPAACVTVTLVDRLVHRAEVIETIEAESYRLKEAKENAT
RTKRRRTKKH"
1998..6263
/gene="epoA"
1998..6263


```
Oy 61 TGCAGCAAAACCCCGCGGCTTCGCGGCTCATGGCGCCC 101
Db 37000 TGCAGCAAAACCCCGCGGCTTCGCGGCTCATGGCGCCC 37040

RESULT 4
LOCUS AR172664 71989 bp DNA linear PAT 17-DEC-2001
DEFINITION Sequence 2 from patent US 6303342.
ACCESSION AR172664
VERSION AR172664.1 GI:17912155
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 71989)
AUTHORS Julien,B., Katz,L., Khosla,C. and Tang,L.
TITLE Recombinant methods and materials for producing epoethliones C and D
JOURNAL Patent: US 6303342-A 2 16-OCT-2001;
FEATURES
source 1. 71989
/organism="unknown"
BASE COUNT 10108 a 23531 c 26617 g 11731 t 2 others
ORIGIN

Query Match 100.0%; Score 101; DB 6; Length 71989;
Best Local Similarity 100.0%; Pred. No. 2,5e-12;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 CGGGATGCCGCTCGCGGCTGTCATGCGCGGCTATCGACGACGGCTGCTGA 60
Db 36940 CGGGATGCCGCTCGCGGCTGTCATGCGCGGCTATCGACGACGGCTGCTGA 36999
Oy 61 TGCAGCAAAACCCCGCGGCTTCGCGGCTCATGGCGCCC 101
Db 37000 TGCAGCAAAACCCCGCGGCTTCGCGGCTCATGGCGCCC 37040

RESULT 5
AF210843 68750 bp DNA linear BCT 21-JAN-2000
LOCUS Sorangium cellulosum strain So ce90 epoethlione biosynthesis gene
DEFINITION cluster, complete sequence.
ACCESSION AF210843
VERSION AF210843.1 GI:6724237
KEYWORDS
SOURCE Polyangium cellulosum.
ORGANISM Polyangium cellulosum.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Myxococcales; Sorangineae; Polyangiaceae; Polyangium.
Molnar,I., Schupp,T., Ono,M., Zirkle,R., Milmanow,M., Cyr,D.D.,
Goriach,J., Mayo,J.M., Hu,A., Goff,S., Schmidt,J. and Ligon,J.M.
TITLE The biosynthetic gene cluster for the microtubule-stabilizing
JOURNAL Agents epoethliones A and B from Sorangium cellulosum So ce90
MEDLINE Chem. Biol. 7 (2), 97-109 (2000)
PUBMED 20130945
REFERENCE 10662695
AUTHORS 2 (bases 1 to 68750)
Molnar,I.
TITLE Direct Submission
JOURNAL Submitted (03-DEC-1999) Natural Product Genetics, Novartis
Agribusiness Research Institute, Inc., 3054 Cornwallis Rd., P.O.Box
12257, Research Triangle Park, NC 27709, USA
FEATURES
source 1. 68750
/organism="Polyangium cellulosum"
/strain="So ce90"
/db_xref="taxon:56"
<1..1826
/note="ORF1"
/codon_start=3

CDS
gene

/transl_table=11
/product="unknown"
/protein_id="AAF26904.1"
/db_xref="GI:6724238"
/translation="ASDALFARATSAVRLDGHGRAPRHVLAERAGIEDLRALREH
LRIQEGPSFHCMLGDLTVELLADQPLASTSFHARRLRHPDMTSDAMLVGDAVLV
RWLAAGAPRLREYEERERARTAOEARRLWLAAPCFADLPFEEDDAGLPLGP
MSPEVAERRLRASYATPELACALMLWLGAGWSGYPAYEMLPENLLVGLFGLPT
ALAAASPGTSEBALRGARLAPSWMEVSSKKSQGNIEALMERLRTVIRAMGNADN
LSRPERAEALIAAEVRLRAOPAPPAAGAGLAVAGVSSGRLSGVLTGDAVSGDGD
LYMPOGRISPVVLLAGNDPFEELAPRISOMLFVAHANAGTISKVLTGESSPLTVARN
QAPMSLVHARCFMAMVNOAMYPDEPRCAFPYVQNSTIWFEEHPPRCLHEPAGSAS
LACDEEHLWCELSAGRIELMRPHRPGAPSRFAVYLGCHPILAAWYPSLTANAHVL
MADPBRRLIYDVKRTGVEPIVLAETRHPPAHVSEDRDIFALTGQPSRDHVEHIR
SCASTVADYOROLMDRPMVLRNRLFFTTMDRLITLARS"
complement(1900..3171)
/note="ORF2: similar to hypothetical proteins from
Mycobacterium tuberculosis and Streptomyces coelicolor and
to Dd-peptidases"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAF26905.1"
/db_xref="GI:6724239"
/translation="MGALISVAPGCGALGAEESQPGODAGALAPAREVMAEVA
ACOMPQAVMLVARGDVDHYDAVTELGSGAPMRDPTFRISMTKAYATVAMVLE
ECKLIDSDVDRLPELARKVLARIDGIDETVPAERITVRDLMFTMGISFEA
SSPIORALDELGLVNAQVPYMTPHGDDEWIRBLGTLPLMHQGAQMMVTSLGVLV
VRADQGRFADFVRRILAPICMROTPDFVPPADKLARFAGCGCYFTDEQGRTRDRD
GASATSPAPFSGAGLVSTVDTLLEFARMLMGVHEGRLLSASVSRMTADHL
TPAQKAASSFFPFGFEETHGWMGVMAVTPADVAEVSEPKYGMDDGEGTSMIDPDEL
TGVNMQASGFLFSGALERFMRSSVYATBSA"
3415..5556
/note="ORF3: similar to Na/H and drug/H antiporters"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAF26906.1"
/db_xref="GI:6724240"
/translation="MHGTFEROVLSTVLTALILVTARSGELARRLROPEVIGELFG
GVVLPSPVYGALAPGFRALROEPANGVGSLISWIGALLILMGITEVDGILKEKA
RGALISALCAIAPPLAAGAFSALVLDRLPBGPLGILYSTANSVAKVILREBSM
RNSVAVOVTLAQVSEVAMVAVMTSSYGSAPLAVARSLALSGFLFVNLVGR
LPHLAMRWVADATRVSKGQVSLVLTFLAALTORGLHPLGLAFALVNLNSAPT
NRPILDGVOTLVAQFAPVFEVLGAKRVVDLSQRLPAAAGTVALLATITAAKVPAA
LGARLGGRLGSEPAALVAVGLNMGGTDLVALVYVGLISMEATVMVAVVALVYTA
SPALLIWEKRRAPPOEESARLEREERARAYIPGERTIVP IVAHAILDGFATDYES
IYASRRKCEYDITELSYEQAPGFSRAGASRLARLRYGIRKQRELRGS
IOAILRASRDHDLVIGARSPARAGMSFGRLQDAIVORAESNVLVVGDPPAAERAS
ARRIIVPIIGLEYSFPAADLAHVALMADEVLVLSAQTDGAVVWRDRERSVRAY
ARSVDEAVFERGRRLGVRVSRVHGVAHPSDEITRELAARAPVLDLVLCGYDHGPIGLRL
YLGSTVESVVSVPVALLVHAGTREGVR"
complement(5612..5992)
/note="ORF4"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAF26907.1"
/db_xref="GI:6724241"
/translation="MDKPIGRTRCAIABSYITGGSNGPEPQNTSHEITACLLASDRDA
QVAIVYFSDRDPAGPYRVTPARRTRHVRNDLIEPEPIPRDTPVASYIESDAPLIV
QHTRLDSRQAEALSLTIAYIDRE"
6226..6675
/note="ORF5"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAF26908.1"
/db_xref="GI:6724242"
/translation="MKHYDTGRFRGRIGHTLGLISMLACGGGSEKTVQGTRLAP
GADARYTADVDDAATTRCLAVYVHLSPERLEASGERPVRVQRPSPRMRVGLD
YNADSRGRLAETIVYANFELLITRAKSSQSPSPSAVIVPTSGV"
7610..11875
```

CDS

/gene="epoa"
7610..11875
/gene="epoa"
/note="Epos a"
/codon_start=1
/transl_table=11
/product="polyketide synthase"
/protein_id="AAF26919.1"
/db_xref="GI:6724253"
/translation="MADPTEERAEDPDIATVGASCRPGVIDSGFWLLEGRDVI
GRVPAERMDAAMPDPPDAPGKTPVTRASTLSVACFDASFEIGISPREALRMPAHR
LLEVEWEALENAIAPSAVAVGTETGVFVIGPEYPAALPQATASAEIDAHGIGTM
PSVAGRIETVALGIRPCVAVDTETVSSSLVAVHACQSLRSGECSTLAGVGLMLP
STVWLKSTRTALARDGRCAKFAADGREGGCAVAVVLRKSGARADGDDILLVIGP
SAIHNDGASGIGTPNGSSOEIVLKRALADAGCAASVYGEAHGTGTGDDPLEIOA
LNAVYGGRRVATPILIGSKYKTLNGHPEYASGITGLKLVLSLOHGOI PAHLHQAALN
PRISMGDLRLVTRAKTPWDMWTPRRAGVSSFEWSTNAHVLEEAATCTPAPAE
RPAPLLVLSARTASALDAQARLRDHLETYPSOCLGDVARSATTBSMEHRLVAAAT
SRERIALDAALAAOGTSPGAVRSIADSSRGKLAFLFTGOGAOTLGMRGLYDWSAF
REAFDLCVRLFNOLDRPLREVMWAEPASVDALDQTAFTOPALFTFEYVALALMS
WGYPELVAGHSIGELVACVAGVFSLDAVETVYAAARGRLMOALPAGAGVSTAPRA
DYAAAVPHASVSIANAVPDQVYVINGAQPVHAITAAAMARCAKTKALHVSATFS
PLAAMPLEAFGRVAESVYRPSIVLVSNLGRKCTDEVSPGIVWHAREVAFAG
VKALHAGAGTFEVEGPKSTLLGLVPACMDPARPALASRAAGDEBATVLEALGGM
AVGLVSMAGLTFPSGGRRPVLPYPMQERERWIDTKADDAAGDRAPGAGHEDVEEG
GAVRGDRRSARLDPPESGRREKVEAAGRPRLDEIDEGVLDHVLRYTERRAG
LGEYETVDAAGLSFNDVOLALGVPPDLGKPKPRLILGEGCGRVYVANGVNGLY
VGQPVILASGAFRTHTTSAALVLPPOALSALAEAAAMPVATVLTATVADRILRDP
GERVLIHAATGGVGLAANVQAQHVAGAVHATAGPERAVLESIGVYVSDSRDRV
ADVAAWAGGEGVDVNLNSGELIDKSFNLRSGRVELESIGVYVSDSRDRV
NLSESLVDKMMLEPARVALLLEELIGLAAGVFPPTATPIARVADAFSMAG
AOHLGKLVLTGDEVOIRIPTHAGAGPSTGDRDLRLRLASAAAPAAALAEARLQ
VSOVLRPEIKVGAELFTRLGNDISLMAVELRNIEATIKLSTFTLSLSPNALLA
ONLLDALATALSERVAENLRAGVQNDPVSAGADQMEITALL"
7640..8920
/gene="epoa"
/note="Region: beta-ketoacyl synthase/malonyl-CoA
decarboxylase"
9236..10201
/gene="epoa"
/note="Region: acetate acyltransferase"
10529..11428
/gene="epoa"
/note="Region: enoyl reductase"
11549..11764
/gene="epoa"
/note="Region: acyl carrier protein"
11872..16104
/gene="epop"
11872..16104
/gene="epop"
/note="Epos P, contains condensation/heterocyclization,
cysteine adenylation, putative NAD(P)H oxidase, and
peptidyl carrier domains"
/codon_start=1
/transl_table=11
/product="nonribosomal peptide synthetase"
/protein_id="AAF26925.1"
/db_xref="GI:6724259"
/translation="MTINQINLEHOGIKLAADGERLQIOAPKNALPNLARISEH
KSTILTMRLORLPAESIVPAPAEHNAPPLTDIOESRTCEAFYVPSGIIHAREVD
CTDDVRLSRAPKVVYARHMDLRAHTLIDTTRKSEORAKRQIOEDLWAMSHCDVS
RIVSIRDAWMSHRIYDTPREPLHYVAVRLDERDORVLTSLDILNVDGSLTIIFKMDL
SEYDEPETSIPVLELSRDYVVALESKKSEBAHORSMDYKRIIAELPPPTLPKMDL
PSTLKEIRFHTEDWLPDSMGRUKRRVGERGLPTGVIIAASEVIGRMASASRFTL
NITLFRNLPVHPRVNDITSDTSWMLDITTRKSEORAKRQIOEDLWAMSHCDVS
GIEVOREAARVGLIORGALPPVLTSLALNDQVGVTSLRLGTFVYSTQPOLDLH
OLYHEDGDLVAMDIYDVGFPPDLIDMLDAVYVLTSLREEBGEVORSLPAOLE
ARASANTNALISEHTLHGLFAARVLEGGAAVPIADALFAERHTYLLDHGEGKLY
EGCARNTIYAVVMEKGEVYAVVAVLEGGAAVPIADALFAERHTYLLDHGEGKLY
LTOPWLDGKLSMPGIGIORLVSEAGVGGDDQPPMIPQTPSDIAYVYVSGSTGLPK
GVMIHDRGAVNTIILDINEREIIPGQVNLASLSLPSLVIVFGIILAAAGTIIVPAP
SKLRDPAHMAELIEREKVYVWNSVPALMRMLVEHFEGRPDSIARSLRLSLSGDWIIV

GLPGLQAIRPGVSVISLGATEASINSIGYPVRNVDLSMASIPYGRPLRNOGFHVD
EALPRPVWVGOLYIGGVGALGYMDDEEKTGRSLVHPETGERLCTGLGYLPD
GNIFPMGEDNOIKIRGVRVGLGLEIFELIKSHRPVAVRPAVGVNDANKLLAYVP
EGTRRRAEODASIKTERIDARAHAAPADGSLDERORFKIARGLRDGKRPVNL
TGDPREAGLDVYIARRSVRFFLEAPLPEVFGRLSLSSVEPDGATLPKFRPSAG
STYPVOYAAVYKSGRIEGVDEGFYHPEFERRLLKLSDHGERAHVRONDVDENA
FNLLFVRDIAIESLYSSSRFCILFAGYMAOLMEQAPCNGVPGQFNDEQVR
PVLDRSHDYYVHGMIGRVDPQFOYCTIGDQSSPRATVRCAPQRFADMLRD
FLRKLPEYWPVYVLELADPLTISNCKVDRKALREKRDSSPRHSGHTAPRALREI
LVAVRYELGLEVYVGLQSFYDLCTSIHYRMSLLQKRLREIATLTELFOYNLGS
LASLRDRSDRLDQRPNMQDVEVRRGRRS"
16251..21749
/gene="epob"
16251..21749
/gene="epob"
/note="Epos b"
/codon_start=1
/transl_table=11
/product="polyketide synthase"
/protein_id="AAF26920.1"
/db_xref="GI:6724254"

Query Match 98.4%; Score 99.4; DB 1; Length 68750;
Best Local Similarity 99.0%; Pred. No. 5.4e-12;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

gene

CDS

1 CGGGATCCCGCTCCGCGCGGCGTTCATGCGCGCGGATCTTGACAGACGGCGTGTGCA 60
|||||
42552 CGGGATCCCGCTCCGCGCGGCGTTCATGCGCGCGGATCTTGACAGACGGCGTGTGCA 42611
|||||

61 TGCAGCAAAACCCCGCGGCGGCGTTCATGCGCGCGGATCTTGACAGACGGCGTGTGCA 101
|||||

42612 TGCAGCAAAACCCCGCGGCGGCGTTCATGCGCGCGGATCTTGACAGACGGCGTGTGCA 42652

RESULT 6
AR193029 68750 bp DNA linear PAT 20-APR-2002
LOCUS AR193029
DEFINITION Sequence 1 from patent US 6346404.
ACCESSION AR193029
VERSION AR193029.1 GI:20238994
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ilyon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and
Gorlach,J.
TITLE Genes for the biosynthesis of epothonones
JOURNAL Patent: US 6346404-A 1 12-FEB-2002;
FEATURES
Location/Qualifiers
1..68750
/organism="unknown"
BASE COUNT 9596 a 22456 c 25539 g 11159 t
ORIGIN

Query Match 98.4%; Score 99.4; DB 6; Length 68750;
Best Local Similarity 99.0%; Pred. No. 5.4e-12;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 CGGGATCCCGCTCCGCGCGGCGTTCATGCGCGCGGATCTTGACAGACGGCGTGTGCA 60
|||||
42552 CGGGATCCCGCTCCGCGCGGCGTTCATGCGCGCGGATCTTGACAGACGGCGTGTGCA 42611
|||||

61 TGCAGCAAAACCCCGCGGCGGCGTTCATGCGCGCGGATCTTGACAGACGGCGTGTGCA 101
|||||

42612 TGCAGCAAAACCCCGCGGCGGCGTTCATGCGCGCGGATCTTGACAGACGGCGTGTGCA 42652

RESULT 7
AR199551 68750 bp DNA linear PAT 20-APR-2002
LOCUS AR199551
DEFINITION Sequence 1 from patent US 6355457.

ACCESSION AR199551
VERSION AR199551.1 GI:20249625
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and Goriach,J.
TITLE Genes for the biosynthesis of epothiliones
JOURNAL Patent: US 6355457-A 1 12-MAR-2002;
FEATURES
source 1..68750
/organism="unknown"
BASE COUNT 9596 a 22456 c 25539 g 11159 t
ORIGIN

Query Match 98.4%; Score 99.4; DB 6; Length 68750;
Best Local Similarity 99.0%; Pred. No. 5.4e-12;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGGGATGCCGCTCCGGCGCTGCTCATGCGCGCGATCCTGAGACGAGGGCTGCTGA 60
|||||
Db 42552 CGGGATGCCGCTCCGGCGCTGCTCATGCGCGCGATCCTGAGACGAGGGCTGCTGA 42611
|||||

QY 61 TGCAGCAAAACCCCGCGGCTTCGCGCGCTCATGCGCGCC 101
|||||
Db 42612 TGCAGCAAAACCCCGCGGCTTCGCGCGCTCATGCGCGCC 42652
|||||

RESULT 9
AR199567
LOCUS AR199567 68750 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 1 from patent US 6355459.
ACCESSION AR199567
VERSION AR199567.1 GI:20249641
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.

REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and Goriach,J.
TITLE Genes for the biosynthesis of epothiliones
JOURNAL Patent: US 6355459-A 1 12-MAR-2002;
FEATURES
source 1..68750
/organism="unknown"
BASE COUNT 9596 a 22456 c 25539 g 11159 t
ORIGIN

Query Match 98.4%; Score 99.4; DB 6; Length 68750;
Best Local Similarity 99.0%; Pred. No. 5.4e-12;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGGGATGCCGCTCCGGCGCTGCTCATGCGCGCGATCCTGAGACGAGGGCTGCTGA 60
|||||
Db 42552 CGGGATGCCGCTCCGGCGCTGCTCATGCGCGCGATCCTGAGACGAGGGCTGCTGA 42611
|||||

QY 61 TGCAGCAAAACCCCGCGGCTTCGCGCGCTCATGCGCGCC 101
|||||
Db 42612 TGCAGCAAAACCCCGCGGCTTCGCGCGCTCATGCGCGCC 42652
|||||

RESULT 10
AR201097
LOCUS AR201097 68750 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 1 from patent US 6358719.
ACCESSION AR201097
VERSION AR201097.1 GI:20251985
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and Goriach,J.
TITLE Genes for the biosynthesis of epothiliones
JOURNAL Patent: US 6358719-A 1 19-MAR-2002;
FEATURES
source 1..68750
/organism="unknown"
BASE COUNT 9596 a 22456 c 25539 g 11159 t
ORIGIN

Query Match 98.4%; Score 99.4; DB 6; Length 68750;
Best Local Similarity 99.0%; Pred. No. 5.4e-12;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGGGATGCCGCTCCGGCGCTGCTCATGCGCGCGATCCTGAGACGAGGGCTGCTGA 60
|||||
Db 42552 CGGGATGCCGCTCCGGCGCTGCTCATGCGCGCGATCCTGAGACGAGGGCTGCTGA 42611
|||||

QY 61 TGCAGCAAAACCCCGCGGCTTCGCGCGCTCATGCGCGCC 101
|||||
Db 42612 TGCAGCAAAACCCCGCGGCTTCGCGCGCTCATGCGCGCC 42652
|||||

RESULT 11
AR208671
LOCUS AR208671 68750 bp DNA linear PAT 20-JUN-2002
DEFINITION Sequence 1 from patent US 6383787.
ACCESSION AR208671
VERSION AR208671.1 GI:21509886
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and Goriach,J.
TITLE Genes for the biosynthesis of epothiliones

JOURNAL Patent: US 63873787-A 1 07-MAY-2002;
FEATURES Location/Qualifiers
SOURCE 1..68750
/organism="unknown"
BASE COUNT 9596 a 22456 c 25359 g 11159 t
ORIGIN
Query Match 98.4%; Score 99.4; DB 6; Length 68750;
Best Local Similarity 99.0%; Pred. No. 5.4e-12;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 CGGGGATCCCGCTCCGCGCGCTTCATGCGCGCGATCTGTGACACGCGGCTGTGA 60
|||||
DB 42552 CGGGGATCCCGCTCCGCGCGCTTCATGCGCGCGATCTGTGACACGCGGCTGTGA 42611
|||||
QY 61 TGCAGCAAAACCCCGCGGCTTCGCGCGCGATGCGGCC 101
|||||
DB 42612 TGCAGCAAAACCCCGCGGCTTCGCGCGCGATGCGGCC 42652
|||||
RESULT 12
AX024384 16124 bp DNA linear BCT 15-SEP-2000
LOCUS
DEFINITION Sequence 82 from Patent DE19846493.
ACCESSION AX024384
VERSION AX024384.1 GI:10184588
KEYWORDS
SOURCE Polyangium cellulosum.
ORGANISM Polyangium cellulosum
Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
Myxococcales; Sorangineae; Polyangiaceae; Polyangium.
REFERENCE 1 (bases 1 to 16124)
AUTHORS Beyer, S. and Mueller, R.J.
JOURNAL Patent: DE 19846493-A 13-APR-2000;
BIOLOGICAL FORSCHUNG GMBH (DE)
FEATURES Location/Qualifiers
1..16124
/organism="Polyangium cellulosum"
/db_xref="taxon:56"
BASE COUNT 2336 a 5400 c 5695 g 2693 t
ORIGIN
Query Match 75.2%; Score 76; DB 1; Length 16124;
Best Local Similarity 85.0%; Pred. No. 6.4e-07;
Matches 85; Conservative 0; Mismatches 15; Indels 0; Gaps 0;
QY 1 CGGGGATCCCGCTCCGCGCGCTTCATGCGCGCGATCTGTGACACGCGGCTGTGA 60
|||||
DB 5317 CGGGGATCCCGCTCCGCGCGCTTCATGCGCGCGATCTGTGACACGCGGCTGTGA 5376
|||||
QY 61 TGCAGCAAAACCCCGCGGCTTCGCGCGCGATGCGGCC 100
|||||
DB 5377 TGCAGCAAACTCCGCGCGCTTCGCGCGATGCGGCC 5416
|||||
RESULT 13
AX024277 16124 bp DNA linear PAT 15-SEP-2000
LOCUS
DEFINITION Sequence 82 from Patent DE19846493.
ACCESSION AX024277
VERSION AX024277.1 GI:10184551
KEYWORDS
SOURCE Polyangium cellulosum.
ORGANISM Polyangium cellulosum
Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
Myxococcales; Sorangineae; Polyangiaceae; Polyangium.
REFERENCE 1 (bases 1 to 16124)
AUTHORS Beyer, S. and Mueller, R.J.
JOURNAL Patent: DE 19846493-A 82.13-APR-2000;
BIOLOGICAL FORSCHUNG GMBH (DE)
FEATURES Location/Qualifiers
1..16124
/organism="Polyangium cellulosum"

/db_xref="taxon:56"
BASE COUNT 2336 a 5400 c 5695 g 2693 t
ORIGIN
Query Match 75.2%; Score 76; DB 6; Length 16124;
Best Local Similarity 85.0%; Pred. No. 6.4e-07;
Matches 85; Conservative 0; Mismatches 15; Indels 0; Gaps 0;
QY 1 CGGGGATCCCGCTCCGCGCGCTTCATGCGCGCGATCTGTGACACGCGGCTGTGA 60
|||||
DB 5317 CGGGGATCCCGCTCCGCGCGCTTCATGCGCGCGATCTGTGACACGCGGCTGTGA 5376
|||||
QY 61 TGCAGCAAAACCCCGCGGCTTCGCGCGCGATGCGGCC 100
|||||
DB 5377 TGCAGCAAACTCCGCGCGCTTCGCGCGATGCGGCC 5416
|||||
RESULT 14
AF319998/c 49736 bp DNA linear BCT 28-MAY-2001
LOCUS
DEFINITION Stigmatella aurantiaca myxalamid biosynthetic gene cluster,
complete sequence.
ACCESSION AF319998
VERSION AF319998.1 GI:14210834
KEYWORDS
SOURCE Stigmatella aurantiaca.
ORGANISM Stigmatella aurantiaca
Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
Myxococcales; Cytophactereae; Cytophactereae; Stigmatella.
REFERENCE 1 (bases 1 to 49736)
AUTHORS Slakowski, B., Nordstieck, G., Bloeker, H. and Mueller, R.
JOURNAL Novel features in a combined polyketide synthase/non-ribosomal
peptide synthetase: the myxalamid biosynthetic gene cluster of the
myxobacterium Stigmatella aurantiaca Sgals
Chem. Biol. 8 (1), 59-69 (2001)
MEDLINE 21110452
PUBMED 11182319
REFERENCE 2 (bases 1 to 49736)
AUTHORS Slakowski, B., Nordstieck, G., Bloeker, H. and Mueller, R.
JOURNAL Direct Submission
TITLE Submitted (07-NOV-2000) MX, GBF, Mascheroderweg 1, Braunschweig
JOURNAL 38124, Germany
FEATURES Location/Qualifiers
1..49736
/organism="Stigmatella aurantiaca"
/db_xref="taxon:41"
complement(2..1306)
/note="ORF1"
/codon_start=1
/transl_table=1
/product="putative aspartokinase"
/protein_id="AAK57182.1"
/db_xref="GI:14210835"
/translation="MRPGSARQGRGTSAALIVOKYSGTSYDTERIKNAKRCIA
AOKAGHDVVVVVSMGSETNRLIKVIOIRPERODVVVAVAGVSGVLAALIO
TOKRAVSPFGHOVIRITDSTFSKARIKSIDARINYADLKKNNIVAVAGGODEIGN
VTTLGRGSGDTTAVAAALAKADCEITTYVDGYTTDPNNVPAARLEIATYEMHE
LASVAGAVVQIRSVFEMKRYKVPDLMWVSFTDDGVTVCEADAMEVDVLSGLAYDKN
EAKIAGVGVDPVGAIAKIGALDEOSIYVDLIVONYSKGRDVTFTVVKADLTAK
DVYKIAKSVKAESEVDQSVISYIVGVGMNRHSGVAAAFVLAGAGVNVOMISTS
EIKVSCVHSNTELAIVRSHTAFGLDKPAGAVSEVSLKG"
1363..2895
/note="ORF2"
/codon_start=1
/transl_table=1
/product="putative histidine ammonium lyase"
/protein_id="AAK57183.1"
/db_xref="GI:14210836"
/translation="MSRRPLNIDSDITKLEETIIQVARRHTVVELAPAAARVAAARDL
VDRVAAGDPTSYGINTGETGLAEYRDKKDLREIQRNLLISHAGVSGPLDEPARVL
LLLRNCVNLAKYSGIRPETLALALEMLNRDVPVPERGVSAGSDIAPLAHALVETI
GESEAFYKGERLPAALALERAGLKPVYLEKEGIALVNGQAMCAVGTLLQALAEMLA
DLADLGAAMTLEGLTSGHKPFPEIDGVRAHEQKCAALRLRELDLSALVESHVNS

```

gene
CDS
    complement(3275..7822)
    /gene="mxaA"
    complement(3275..7822)
    /note="non-ribosomal peptide synthetase: contains a
    condensing, adenylation, and PCP reductase domains"
    /codon_start=1
    /transl_table=11
    /product="MxaA"
    /protein_id="AAK57184.1"
    /db_xref="GI:14210837"
    /translation="MSVRELLTTLIAKGVRLTEGDKLOVEASKAVLPBELBAELSAN
    KADIIAFLKHSRRETSGLCOPLARPPVITLSSQGEKLMFLDRLSPGTSQYNHLG
    LKARGALDTRALRSLDELVRHREVLRTHEPVEGSPQIVSPDTGVALITVELGL
    TESQREEEIYRSDLEAOQPDPLAKGPLRLVLAIGPDPALFVTKHHIITDGSLG
    VVRELIMLYASFVGGQAALPPVSKQFDSALREBAMLAGSGARERAYWKGKLG
    PLLOLPVDAVSTSHRGATVPALSPALDRELIRASREGCSFEPTVFAAFSALIH
    RSGVDGRCGVIANRGSVPADILGFIANTLALRCDLSGEPTESQMLARAKVYLEA
    LDYQALPSEVQIVGASRDGGLNPLVRACTFLESTIPATLDPGTSMSEFLGADGS
    VSGVAKFELSILIASSEKGLACMLESRVFPDASTVERMWGHRVOYLESIVAHPEVPL
    SKPLITAEERGLINDMNGPVLDPALCMHBLVOAQERTPOAAVAVSGQRTLYAE
    LNRANQLAHLRLRIGIKKEERGLDVERTEDIVIGLAILKAGGAYVLDPAYPEK
    LALTEDQVAPVLTOOQLRVPRLPATQAVVCLDKDMPITGAESDTPRITRAVIA
    YLITSGSGTKRKVMIEHRNVAFLIAMSVPFSKRELAGTIASTISICDTSVPELFT
    PLCCAKYIVAKNALEPELPAPARETTLINIVPSAGALLRSGGPSSAIVNLGEEA
    LAGALVDQIYQLDHRVDVNLGPSETTYTSTPTNRNGQPTTIGPNSVTLNADLS
    NREMPIGVEVYIGGMVARGYLGREPLTAERVRSPFGSGSEARLYRTDGLRML
    PDGOLEYLRMDQVKLRGRIEELGICAVLMEHSGIRDAVVVREGLADQOLVAY
    VGRGEKAPPELRDYIKSLKPEYVWPELFGIDLPLTPNGKVDRAALPAPERHSG
    PAKHEVAPTPGEESLAAITMROVLGVEOGADNPEELGSHLLYRVLVARSASGA
    DIPRLALDAPLLEEMARAVEAKTGSVDHADVMEADAVLDAEIALGKRLPPTG
    ALRTLLTGATGFLAEELCEKRTDARTYCLVRSKTEQEGEMNRIKRLNEISLWNE
    ALAPRIVPRGDIQPLGLISEKFORSEEDAIYHNGALNPLLYPYMSAAVNLG
    TREIRLATRTRIKPLHYSTVSLPLGRKAIREDPELGESSLVGGAQSKWAEK
    LYREASRGLPVTILRPGRVGTHSRGTAMTDVLCRTLKGCVRMGVAPSDALLDT
    PVDVSAIVDLSMRPESIGORTYLVNPGVRADMMNMRAPGYGLRVLPLVDMLSE
    LGSAAASDSELDLMLFLAOQVPPEDRSVSGPPMYVCDSDTTLKALGGTSCPSVDAS
    LSTYLSLVHGFLLKAPVR"
    complement(7819..11520)
    /gene="mxaB1"
    complement(7819..11520)
    /note="PKS: ketosynthase; acyltransferase"
    /codon_start=1
    /transl_table=11
    /product="MxaB1"
    /protein_id="AAK57185.1"
    /db_xref="GI:14210838"
    /translation="MRAGNGHPLILGVSTTSVQENRWFESALSKKPSYIADHRVG
    GAVVYPGAYILEMALSAARELSGEKAYELVDTSFKEGLLPGEQEVWMAISGEGD
    ALSTQLDSQKATSEPNAPAGMAHVVGQIRTPTEGGSTQTPSPFEAIRKCEVIAQA
    AHYEELASGVAYGPAFOVQOIMRGRAEALGHVRLPEPLARASRYRHPALLACF
    QLVAAVAPQEDLTPAGPAPVPAALHRLRHEOGTEVFCVHLRPSSEGRKRVYEDL
    VLRDASGRILEVALGIRVOOLADPAAEHNGKDLFPALERHNASAVOQVAPGASGR
    WLLINDSEIADYSQLEBERGEVIRYDLAAGGRYVDATSTQAPFGVYSSEAFGA
    GSAPRGVILHTGLEPREVDPLMGCGVHLVVALSRVSLASPPRLMLVTRGVHHSGS
    RSASOVQASLWGLGRTLAHESELKCTRVDSASMTGDDAALMARELLVADDEEI
    SLRADGRHYGRIVRSGPRAPSEAVYPAGRGPHLEMDQGRLEFRARPPGEGEL
    EIEVEVALGPGEGTGPPEMGAGEGICGSRVVAIGEVKNVAVGEERILVAGSL
    GSHVSPVAGCTAIRSSSHGGBAALGAANTVLPAMYALPHGLERKGRVAGTASGL
    GRAAIKLAKRAGAEVFAVATTDQRASLRELCADYVMDSDSPYSRVLDTGCGKVE
    LAIVPAGVAEELERSLPTLSEGARVLDLARSASARTGDANVAMCVADTGVHNGER
    FASLQEVKAEAEAGCLPSAAADAKALIKDKARVYVPAADARLRADGTLVNGEL
    GGLGLAVAKMVEOGAGHLVLYGRDVTLSPEQEEVNTALEAGARVARRAVYSRAQ
    IARILSEIAEGGAPLRGVYHAGVLDGVLTOQTVERRRVAAPVYIGVWNLVJLRE
    APLDFEVLKSSAASMEGAPGGSNVAAGNAFLDALAHHRRALGCLPGLSTIMGPFSEVL
    AAAQNRGERLQAGSDSLTPAEGNTILGRLLDGDQVQMAVPLEKRWVEYTPRKS
    SPWLSELVPADAGDEVGEARVALLTLETATVPQAKAALQEFVREQLARLYLDSAR

```

```

gene
CDS
    complement(11487..14303)
    /gene="mxaB2"
    complement(11487..14303)
    /gene="mxaB2"
    /note="PKS: dehydratase; enoylreductase; ketoreductase;
    ACP"
    /codon_start=1
    /transl_table=11
    /product="MxaB2"
    /protein_id="AAK57186.1"
    /db_xref="GI:14210839"
    /translation="MSKQEPAPQPSALORALLVKRNQARLDAVERARTETAVTGL
    GCRPFGAVIDSSYWKILRDGDVALREVPESWDVPGHEDSRVPGVMYGRGGFLD
    DYEHFDLAEFFGISPRESALIDPOQRLVLEAMEALENGLAPDLAGSKTQVGMVMS
    SYMARLAKENDATPFDGMATNGNYSYTPGVSGVSLDLCQCFMVPDTRACSSLYALH
    LASESLRGEESMLAAGVNLILSPETMCLCSQALASDCKTFDASADGVNARGC
    CGVILKRLSDQRDGDILALIRGSANHDSASGGLVPPNPGSOQAAVORALDMARI
    APALVGYLEAHGTGTPLDGPIELRALGAVLGRPADRPFPGSVKTMIGHLEPAGI
    AGVITKTIISLOKEIIPHLHFTPPHVEMDRIIPARVPERVPMFAHGRRJAGVSF
    GLSGINAHVYLEAPASTAPTAFAEDKRAOLIVYLSAKTPRALSSIAEAMSSPTKMEG
    GSLADISTASLSTRPHERIALAKSKQELIEQLEARIACKSHAGLSGSRGRTLGO
    KIVFVPGGSGQLMGQRLAENVTFEADIDWCHIASRTDMSLRDLTPGQQR
    NMDIDVIOPTLEALQVLAALAAVWRMLGLEPFAVAVGSMGEVAAAHVAGALSLEDAARI
    CERSKILRVSQGAMAVVDLSREQAEVIRQAEEDRIAIASNGPSTVLSGDPALK
    EYLERLERREYRCRWKVDVASHSQMDPLRODILDRMAALIPGSAVPLXSTVYAG
    IDGRGLVAPYWRNLROPVLPADYORLIODGNVYFTELSPPRLVLPVDMALRGDT
    SGRGLVPSLITDQDEMPSLSSVCAALTTRVIAWIGSGIPIPSVAGLFLRTRGASG
    TMSQOQLRRRSCALAREVIPS"
    complement(14300..29878)
    /gene="mxaC"
    complement(14300..29878)
    /note="PKS: modules 1,2 and 3 contain KS, AT, DH, KR, and
    ACP regions"
    /codon_start=1
    /transl_table=11
    /product="MxaC"
    /protein_id="AAK57187.1"
    /db_xref="GI:14210840"
    /translation="MANRNDESNEISPAKRMILVALEKQTRLNAVEGAKEPIATVGM
    ACRPQADNVEAFMRLLDGGVDAREIDLEPRKDVEDADIGGAKRAGWGFSDVSFD
    PGFEGISPREAVRMDPOORLLLEVAMELEIDMGDPOVLDLSSGSRVFTGACNDYDHCM
    QYERPETGDAFSATGVAASVLSGSLSYFTNLGGPLRVKATDACSSTLGGCAVSLNR
    RECNMAGAGVNLILSPOSVLVSRLKLOALSPGRKAPDASNGSTREGGCIVLYKR
    LSDALADGHILATIRGSAIINDQGSTGILTPNVLSOALIRLOALESSGIAEOVSY
    EAHGTGTPLDGPIEKALRETYGVPRPGSGVCGTGSYTNNGHLESNAAGIANKVYL
    AAKHOSITPPLHLKQVNPRIQLESALTIPTKLTWKASGQRRRAGVSSFGISGNH
    VIIEBAPAVAKDALDQREEVPLRSARSEALVALAERYAEKLRGEGGSLSDVCYTA
    SAHRTHTHRLVALDLSREHTERLARARAGEOPSGANGRAKAFEGRRKIYFVYPGQG
    AORAGMGRLMSESEVPFDALVYKDKALVKGMLVDEIDAEORSLKAEFDOSPCGMVH
    LEFAVEVLTALMRSGVIEPDVAMGSMGEVVAHAHVALSLIEDALIIICHSKMLRTL
    GGRCATMANVELSEVEABEATRPNNGKITHLGSLNRPROYLSGEPAYKEAVEALGRK
    GVFTRWKADVPSSHPLVEVLSIRELGRATIRTPPAADIPLCSTVATGFEUAKIDAA
    HNGNMLRNEVPRAAAOHLLASDHCTFEVSHAPVLPAVEGTMSAGTEGVALASLR
    NREDHRTMLSSIGALYSOGSVLNNRNYPGQGLVPTPTWQEREMWMEBAVAGVR
    ASRTSSGPHLIERIDLSAOGSVSLNNOVLGPKHPEPLLAERHVGCVALPAAAGM
    ALSAKOTPLALHESLIRGDSAVYKAPVADLDELRAKCEBASVEAHYQTLKEIGLE
    TPGSTRVVRKELIRANGELDGRVQGEGLTAEERI SLHPALDSCFOVLAALBEGLDS
    BEGVYVPALSLRLVLRPEALMHSVITISQAEAGST IDECVLILNDQORVAAEV
    RLVCKRIIAASAPADEVIRWITSAQEMQOIRGARSGETSSRTAGMVLFPADARIGOS
    LASGIAERGGCGVVEEAGSVYRLEADRYRINPSKLEDFOLALKEAFDOSPCGMVH
    LYGLDVA NVMPRESEFEALSTISSTVYILQAMRQCMRQMPRLMTVTVQVYIASC
    EOPGSPAAPLVGLAAVYTTHEPELRCHAHVDSQOPSDEBLASTLEILLGGAEDRIA
    LRGAVRHAARLEKRGISGRKAAPREFTADGYLLITGGLGGLLEVAKNF IORGAGNLA
    LTRGAPSEABEAGKIGSELRANGAEVRYRADVTSADGVARILGS IDSELPLRGVHA
    AGVINDGYLTODEPRRLAVLAPKVOGAMNHLVQOQGLPELFEVSSAAASMGSPGO
    GNYAANAMPMDLAAHRRRAAGIPGLSSGMHPADGVLAAVTINRCDRLALGVMGMP
    AQALEAIBRLVQEAQVAVYVDLRQMRREPLSAAGSPSLYSIMQEKVSTPASKGS
    ARELILAAPEBQRKGIKLSYLEQGLRLIKREPSITFEDQPGDGLGMSLGLLELRNK
    IESGIKTLTSLATLIYAPTVQALLTYYLKGLEPYLEAEARPTVREPTQTEPRVSGVAP

```

Query Match 52.5%; Score 53; DB 1; Length 49736;
Best local similarity 73.1%; Pred. No. 0.033;
Matches 68; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

Qy 9 CCGCTCCGCGCGGCTGTCATGCGCCGATCTGAGACGAGCGGCTGCTGATGACGAA 68
Db 8739 CCGGTGAGAGCGGTATCATGACGCGCGGGGTGTGTGAGACGAGGCTCTGTTGACGAC 8680

Qy 69 ACCCGCGCGGCTGTCATGCGCGCC 101
Db 8679 ACGGTGAGCGGTGTCGCGCGGTGATGCGCC 8647

RESULT 15
AF232752 3225 bp DNA linear BCT 15-FEB-2002
DEFINITION Mycobacterium avium subsp. paratuberculosis polyketide synthase 7
ACCESSION AF232752
VERSION AF232752
KEYWORDS AF232752.1 GI:8927424
SOURCE Mycobacterium avium subsp. paratuberculosis.
ORGANISM Mycobacterium avium subsp. paratuberculosis
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacterineae; Mycobacteriaceae;
Mycobacterium; Mycobacterium avium complex (MAC).
1 (bases 1 to 3225)
REFERENCE Bannantine, J.P. and Stabel, J.R.
AUTHORS Identification of two Mycobacterium avium subspecies
TITLE paratuberculosis gene products differentially recognised by sera
JOURNAL from rabbits immunised with live mycobacteria but not heat-killed
MEDLINE mycobacteria
21432816
PUBMED J. Med. Microbiol. 50 (9), 795-804 (2001)
11549181
REFERENCE 2 (bases 1 to 3225)
AUTHORS Bannantine, J.P.
TITLE Direct Submission
JOURNAL Submitted (09-FEB-2000) Bacterial Diseases of Livestock, National
Animal Disease Center, 2300 N. Dayton Ave., Ames, IA 50010, USA
FEATURES
source
1..3225
/organism="Mycobacterium avium subsp. paratuberculosis"
/sub_species="paratuberculosis"
/db_xref="taxon:1770"
<1..2038
/note="Pk57"
/codon_start=2
/transl_table=11
/product="polyketide synthase 7"
/protein_id="AAF82076.1"
/db_xref="GI:8927425"
/translation="DORLTAVPPGLSLVAAGVVFVLTALYGLSVLAGLPGPERVL
VHTATGGVMAVOLARHYGAEVFATASRGKMDTLRAMGEDDDHISDSPDFEOKFL
AATGGAGVDVVLNLSAGEFTDASRLIAPGRTIEMKTDVDPDVAQRGARVRA
FDLMEAGPDRTRAMLAETVGLDQGVLTPLKTFDVRCASATRVISQAHIGRVYL
TVPSGLDEVLSGCGGLAQSTVLTGGGAGSALARLVDYRVGAVVLSRTGAQA
AGAAELVDDELRAQASASVLAQVADRAVATMLAGLPAGYPLRGIVHAAGILDDGL
SSLIPDRDAVIRAKVDGAMNHELTKDLISAFVSSMAGIVTGPQANYAANSF
LDGLVARRADGLAGLSIANGIMEASMTAHLGDRKARMSRTGIAPISPEOALAF
DAAMLVETPVYVARLDRAALSENIALPLRLRLAGPTRRVITTDADYTRSMGLAA
RLHGLSPARRRELVDVCGNAAMVILGPNPADINAGAFODLGPDSLTAVALRNRLK
NATGLTSLPTLIFDYPIPVVLAELHLSRLAGSGDDQDPLMGRENDDITRELQALLGAA
HMNSDDKAVLRTRIHGLGALPAGDPSAPLDDLEAATESQLFALIDELGR"
2058..>3225
/note="Pk58"
/codon_start=1
/transl_table=11
/product="polyketide synthase 8"
/protein_id="AAF82077.1"
/db_xref="GI:8927426"
/translation="MPGTGHLDTLKLRLTADLRTRRRVLEBGRLESPVAVGMACR

YAGVDSPEALMDLVIEGRDITVSDPVDRCWDEGLYDPPDANKMYTRQGSFLQHA
GDFPAGFEIGIPSEALAMDPOORITMLEICWELRABAGIDPSALRGTAQVAYTHAG
YGEVKEGEBEGYGLTSGTSLSTSGRSVYVGLBEPAYSDPACSSIVAMHIAOSIR
SGECDLALAGVTVMAITPAATVRSRQNALPDRCKYRAGAAAGTSSBAGVLYLE
RLGDARRIGHPVMAVLRGSAAVNDGASNGLAPNPGSQORVIRALANAGLSAVDV
VEGHTGTGLDPIEAQALLATYGDPRADRLPLWGISIKNIGHTSAAAGVAYT"
BASE COUNT 419 a 1080 c 1238 g 488 t
ORIGIN

Query Match 47.7%; Score 48.2; DB 1; Length 3225;
Best local similarity 67.3%; Pred. No. 0.65;
Matches 68; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

Qy 1 CCGGATGCGCTCCGCGCGCTGTCATGCGCGGATCTGAGACGAGCGGCTGCTGA 60
Db 945 CCGGCTATCCGCTCGCGGAGATCGTGCACGCGCGGGATCTCGACGAGCGGCTGCT 1004

Qy 61 TGCACMAACCCCGCGGCTTCCGCGGCTGATGCGCGCC 101
Db 1005 CCTGCTGATCCCGGACCGGCTGATGCGGCTGCTGCGG 1045

Search completed: November 6, 2002, 13:29:59
Job time : 593.625 secs


```

/note="Vector: pZero-2 (Invitrogen): Site_1: EcoRV;
Genomic DNA was isolated from stationary phase cells. For
this library, DNA was sheared to give a tight size
distribution of 1-1.5kb fragments, blunt-ended with T4 DNA
polymerase, dephosphorylated with Shrimp Alkaline
Phosphatase and ligated into pZero-2 vector's EcoRV site."
BASE COUNT
ORIGIN
88 a      120 c      160 g      95 t

```

Query Match	32.9%	Score 33.2;	DB 17;	Length 463;
Best Local Similarity	59.6%;	Pred. No. 37;		
Matches 56; Conservative	0;	Mismatches 38;	Indels 0;	Gaps 0;

DY 2 TCGCCGCCACCTCGGCGGGAGATATTTCGACCCTGTACGCCGAGAAGCGGGCT 61
 | | | | | | | | | | | | | |
Db 166 TGGCATCTCTTGCAGAGCGCAGCGATGACCGCTGTCCTGTACATGAGCGCA 222

QY 62 GGCTGCGGAGCAGGGGATCGCGCACGTGATGGA 95
 ||||| ||| ||| | | | | | | |
 Db 226 CGCTGAGCGGGCCACAAGCTGGAAGCTCGCATCGA 259

RESULT 4			
BG836681.			
LOCUS	BG836681	752 bp	mRNA
DEFINITION	Zm08_06a01_A		linear
			EST 25-MAY-2001

ACCESSION	BG836681
VERSION	BG836681.1
COORDINATES	GI:14203004

SOURCE	Zea mays
ORGANISM	Zea mays

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Zea.

REFERENCE	1 (bases 1 to 752)
AUTHORS	Harris, L.J., Balcerzak, M., Allard, S., Saparno, A., Courcoux, P., De

TITLE Expressed Sequence Tags from Developing Maize Kernels Six Days after Silk Channel Inoculation with *Fusarium graminearum*

JOURNAL Unpublished (2001)

COMMENT
Contact: Harris, Linda J.
Eastern Cereal and Oilseed Research Centre
Agriculture and Agri-Food Canada
Bldg. 21, Central Experimental Farm, Ottawa, Ontario, K1A 0C6,
CANADA
Tel: (613) 759-1314
Fax: (613) 759-6566
Email: harris1@em.agr.ca.

FEATURES	Location/Qualifiers
source	1. .752

```

/organism="Zea mays"
/cultivar="C0430"
/db_xref="taxon:4577"
/clone="Zm08_06a01"
/clone_id="Zm08_AAFc_ECORC_Fusarium_graminearum_inoculate
d_corn_ear"
/tissue_type="Developing kernels (silpocrossed)"
/dev_stage="10-11 days post-silk emergence"
/note="Vector: Bluescript SK+/XhoI-ECOR1; Site_1: ECOR1;
Site_2: XhoI; Field-grown maize ears were silk
channel-inoculated in the morning (-10 am) with 1 ml of a
Fusarium graminearum macroconidial suspension (500,000
spores/ml) and whole ears were collected and immediately
frozen in liquid nitrogen 6 days later."
BASE COUNT      173 a      220 c      201 g      143 t      15 others:

```

Query Match	32.5%	Score 32.8	DB 12,	Length 752,
Best Local Similarity	58.0%	Pred. No. 50,		
Matches 58; Conservative	0;	Mismatches 42;	Indels 0;	Gaps 0;

[illegible]

RESULT 5
BE360248

DEFINITION	DG1_62_F06.b1_A002 Dark Grown 1 (DG1) Sorghum bicolor CDNA, mRNA sequence.
ACCESSION	BE360248

KEYWORDS	EST.
SOURCE	Sorghum.
ORGANISM	Sorghum bicolor

ciade; Panicoideae; Andropogoneae; Sorghum. 1 (bases 1 to 580)

TITLE	AN EST database from Sorghum: dark-grown seedlings
FORMAT	flatfile (2000)

COMMENT
Contact: Cordonnier-Pratt MM
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860

Fax: 706 583 0210
Email: impratte@uga.edu
Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20.

Seq primer: JEN REV
High quality sequence stop: 539
POLYA=NO.

FEATURES	Location/qualifiers
source	1. .580

```

/organism="Sorghum bicolor"
/db_xref="taxon:4558"
/clone_1b="Dark Grown 1 (DG1)"
/notes="Organ: 5-day-old dark-grown seedlings; Vector:
lambda Zap; Site_1: XhoI; Site_2: EcoRI. The library was
made from poly-A RNA in the cloning vector lambda Zap II.
Clones to be sequenced were prepared by mass excision."

```

BASE COUNT	104 a	189 c	172 g	115 t
ORIGIN				

Query Match	32.3%	Score 32.6	DB 10	Length 580
Best Local Similarity	58.9%	Pred. NO. 55		
Matches 56	Conservative 0	Mismatches 39	Indels 0	Gaps 0

```
QY      6 CCGCCACCTCGGGGCGAGATATTGCCACCCTGTACGCCGAGAACGGGGTGGCT   65  
        ||| ||| | ||||| | ||||| | ||||| |  
Db     34 CCGTCATCGCAGCGCACTAGCATCGGCACCGCGTGAGTCCGACTTGGCAGCGGACT   93
```

```

2y      66 GCGGAGCAGGGGATCGCGCAGTCATGGACTCGC 100
          ||||| |||  ||  ||  || |||||
Db      94 GCGGAGGAGGAGCCGCGGCGCGCGGTCGCGC 128

```

BQ217222

DEFINITION	AGENCOCURT_7696437 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:6043949 5' , mRNA sequence.
ACCESSION	B0217222
VERSION	B0217222.1 GI:20398622

KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE NIH-MGC http://mgc.nci.nih.gov/
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
Plate: LLM13285 row: m column: 06
High quality sequence start: 40
High quality sequence stop: 180.
Location/Qualifiers
1..1323
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NIH_MGC_92"
/tissue_type="embryonal carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: testis; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; Oligo-dT primed.
Average insert size 2.5 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library." 4 others
BASE COUNT 292 a 325 c 520 g 182 t
ORIGIN
Query Match 32.3%; Score 32.6; DB 14; Length 1323;
Best Local Similarity 60.2%; Pred. No. 60;
Matches 53; Conservative 0; Mismatches 35; Indels 0; Gaps 0;
QY 13 CTCGGCGCGGAGATTTTGGACCGCTGATACCGGAGAGCGGGCTGCTCGCGAG 72
Db 747 CACGGCGGTGTGATGTGTCGGCGCGCGGCGTGGCAGCGCCGCGCTCGCGCTGG 806
QY 73 CAGGGATCGCGACGATGATGACTGCC 100
Db 807 TCGAGGAGGCGCGCGCGAGAACACGC 834
RESULT 7
BF250997 699 bp mRNA linear EST 15-NOV-2001
LOCUS EST118254 Coccidioides immitis spherule cDNA library Coccidioides
DEFINITION Immitis cDNA clone CIAA807 5' sequence, mRNA sequence.
ACCESSION BF250997
VERSION BF250997.1 GI:16931140
KEYWORDS EST.
SOURCE Coccidioides immitis.
ORGANISM Coccidioides immitis.
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
Onygenales; Mitosporic Onygenales; Coccidioides.
REFERENCE 1 (bases 1 to 699)
AUTHORS Gardner, M.J. and Kirkland, T.
JOURNAL Generation of ESTs from Coccidioides immitis spherule cDNA library
COMMENT Unpublished (2000)
Contact: Malcolm J. Gardner
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301 838 3519
Fax: 301 838 0208
Email: gardneretlgr.org.

FEATURES Location/Qualifiers
SOURCE 1..699
/organism="Coccidioides immitis"
/db_xref="taxon:5501"
/clone_lib="CIAA807"
/dev_stage="spherule"
/lab_host="SOLR"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI"
BASE COUNT 136 a 206 c 186 g 171 t
ORIGIN
Query Match 32.1%; Score 32.4; DB 12; Length 699;
Best Local Similarity 68.2%; Pred. No. 63;
Matches 45; Conservative 0; Mismatches 21; Indels 0; Gaps 0;
QY 1 ATGCGCCGACCTCGCGCGGAGATTTTGCAGCGCTGTACGCCGCGAGAGCGGCG 60
Db 98 ATAGCGGTCATGTAGCGCGGAGGCTTTGTACTGTGCGGACTCTCGAGAAACGCGAG 157
QY 61 TTGCGTG 66
Db 158 TTGATG 163
RESULT 8
B78358 403 bp DNA linear GSS 16-JUN-1998
LOCUS T32E17TF TAMU Arabidopsis thaliana genomic clone T32E17, DNA
DEFINITION sequence.
ACCESSION B78358
VERSION B78358.1 GI:2774997
KEYWORDS GSS.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
1 (bases 1 to 403)
REFERENCE Rounsley, S.D., Field, C.E., Bass, S., Linher, K., Linher, K., Golden, K.,
Berry, K., Granger, D., Suh, E., Wible, C., Adams, M.D. and Venter,
J.C.
TITLE A BAC End Sequence Database for Identifying Minimal Overlaps in
Arabidopsis Genomic Sequencing. Update 3
JOURNAL Unpublished (1997)
COMMENT Other_GSSs: T32E17TR
Contact: Steve Rounsley
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: rounsleyetlgr.org
Seq primer: M13-21
Class: BAC ends
High quality sequence stop: 403.
Location/Qualifiers
1..403
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone_lib="T32E17"
/sex="hermaphrodite"
/note="Vector: BelobAC11; Site_1: HindIII; Site_2: HindIII
; Produced by Rod Wing"
BASE COUNT 69 a 123 c 131 g 80 t
ORIGIN
Query Match 31.9%; Score 32.2; DB 17; Length 403;
Best Local Similarity 63.6%; Pred. No. 67;
Matches 49; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

OY	24	GATATTGTGGACCGGTTGGTACGCCGAGAACGGGGCGTCGGCCGAGCAGGGGATCGC	83
Db	173	GAATTGGCGACTCTCGCGAAGTCCGAGAGCGCGCTTGTCAGAGACCGATGACGC	232
OY	84	GCACGTGATGCACTCGC	100
Db	233	GACATCGATCTTCACGC	249
RESULT 9	CNS05PK8		
LOCUS	DEFINITION	1050 bp DNA linear GSS 26-MAY-2000	
REFERENCE	CNS05PK8	Tetraodon nigroviridis genome survey sequence T7 end of clone	
SOURCE	ORGANISM	066N06 of library B from Tetraodon nigroviridis, genomic survey sequence.	
VERSION	AL348065	AL348065.1 GI:8241835	
KEYWORDS	GSS; genome survey sequence.		
ORGANISM	Tetraodon nigroviridis		
REFERENCE	Bernot,A., Fitzames,C., Wincker,P., Brodier,P., Quetier,F., Saurin,W. and Weissenbach,J.		
AUTHORS	Roeest-Crollius,H., Jallion,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fitzames,C., Wincker,P., Brodier,P., Quetier,F., Saurin,W. and Weissenbach,J.		
JOURNAL	Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence		
REFERENCE	Unpublished		
AUTHORS	2 (bases 1 to 1050)		
JOURNAL	Roeest-Crollius,H., Jallion,O., Dasilva,C., Fitzames,C., Fisher,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J.		
REFERENCE	Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis		
AUTHORS	Unpublished		
JOURNAL	3 (bases 1 to 1050)		
REFERENCE	Genoscope.		
AUTHORS	Direct Submission		
JOURNAL	Submitted (12-APR-2000)		
REFERENCE	This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/Tetraodon.		
AUTHORS	Location/Qualifiers		
JOURNAL	1..1050		
REFERENCE	/organism="Tetraodon nigroviridis"		
AUTHORS	/db_xref="taxon:99883"		
JOURNAL	/clone="006N06"		
REFERENCE	/clone_lib="B"		
AUTHORS	/note="Genoscope sequence ID : COAB006D03Ci-end : T7"		
JOURNAL	BASE COUNT	250 a 231 c 259 g 268 t 42 others	
ORIGIN			
Query Match	31.9%	Score 32.2;	DB 17; Length 1050;
Best Local Similarity	57.6%	Pred. No. 75;	
Matches	49; Conservative	0; Mismatches	36; Indels 0; Gaps 0;
OY	3	CGCCCCCACCTCGCGCGAGATTTTGGACACCGCTGTACGCCGAGAAAGCGGCGTG	62
Db	965	CGCGCGCGCGCGCGCGCGGNGNGNGNCAGCAGCAGCTCGCGCGCGCGCGCGCGCNCG	1024
OY	63	GCTGCGGAGCAGCAGGAGATCGCGCAC	87
Db	1025	CCNNCGCGCGCGCGCGCTCGCGC	1049
RESULT 10	BM379457/c		
LOCUS	494 bp mRNA linear EST 16-JAN-2002		

DEFINITION	MEEST505-H02.univ ISUM6 Zea mays cDNA clone MEEST505-H02 3', mRNA sequence.
ACCESSION	BM379457
VERSION	BM379457.1 GI:18178247
KEYWORDS	EST.
SOURCE	Zea mays.
ORGANISM	Zea mays. Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Zea. 1 (bases 1 to 494) Men,T.J., Qiu,F., Guo,L., Ashlock,D.A and Schnable,P.S. Expressed Sequence Tags from B73 Maize: various stages and tissues including seedlings treated with a variety of hormones Unpublished (2001) Contact: Patrick S. Schnable Schnable Laboratory Iowa State University G405 Agronomy, Iowa State University, Ames, IA 50011-1010, USA Tel.: 515-294-0975 Fax: 515-294-2299 Email: schnable@iastate.edu Individual basecall and confidence value were assigned using the phred software(http://www.phrap.org/). Overall sequence quality assessment and vector trimming was conducted using the Lucy software (version 1.16s, http://www.tigr.org/software/lucy). Lucy parameters were set to ensure an overall trimmed quality of 97.5% or better without any vector fragments in the chosen high quality region of each sequence. Low-quality bases between the poly-T and the high-quality region were replaced with N's to serve as spacers using a Perl program (<code>est_process.pl</code>), written by Dr. Hui-Hsien Chou. PCR Primers FORWARD: primer T7-1 (AA TAC GAC TCA CTA TAG) BACKWARD: primer Sp6 (ATT TAG GTG ACA CTA TAG) Seq primer: universal (GTA AAA CGA CGG CCA GT) POLYA-tes.
JOURNAL	
COMMENT	
FEATURES	Location/Qualifiers 1..494 /organism="Zea mays" /contig="B73" /db_xref="taxon:4577" /clone="MEEST505-H02" /clone_1lb="ISUM6" /tissue_type="mixed" /lab_host="DH10B" /note=Vector: pSP167 (4.43 kb); Site_1: EcoRI; Site_2: NotI; Tissue samples were collected and partially pooled prior to RNA extraction. First-strand cDNAs were prepared from 21 individual pools of oligo-dT selected mRNAs by priming with 21 different NotI oligo-dT tag primers (5'-AACGTGAGAAATTCGGCGCCGNNNNNTTTTTTTTTTTT-3'). Distinguishable 'bar code' tags, (N)6, were used for each separate first-strand cDNA synthesis. Hence, these bar code tags can be used to identify the mRNA pool from which a particular cDNA clone was derived. The 'bar code' tags associated with specific tissue sources are: ATMGCG--Germinated seeds and seedlings (1, 2, 8, 11 DAG); ACTGCC--Mixed mature tissues (17, 21, 39, 69, 77 DAG); CAGGC--Kernels (3, 5, 10, 15, 20, 25, 30, DAG); TAACGC--Adventitious roots (65 DAG); CAGCG--Tassels (3-39 cm, 53 and 56 DAG); AGCTAC--Immature ears (0.2-3.0 cm, 53, 56, 59 DAG); TGACCG--Husks (73 DAG); GACAC--Stilks; ATMGCG--unpollinated first ears; CTAGCG--ear shanks; GAAGCG--etiolated seedlings; AGTAG--callus; GTGAC--Cycloheximide-treated callus; GTACC--Anaerobic treated seedlings; CGTCCA--NA (α-Naphthalene acetic acid)-treated seedlings; GATGC--Kinetin-treated seedlings; AAGCG--ACPC (l-aminocyclopropane-l-carboxylic acid seedlings; CTAGC--ABA (Absciscic acid)-treated seedlings; TAGCG--GA (gibberellic acid)-treated seedlings; GAGGA--JA (jasmonic acid)-treated seedlings; Equal

[illegible]

ACCESSION	Sheared DNA-34021, DNA sequence.					
VERSION	AQ940375					
KEYWORDS	AQ940375.1 GI:6763640					
SOURCE	GSS.					
ORGANISM	<i>Trypanosoma brucei</i> .					
	<i>Trypanosoma brucei</i>					
	Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;					
REFERENCE						
AUTHORS	1 (bases 1 to 640)					
	El-Sayed,N., Zhao,S., Zhao,H., Gill,S., Suh,E., Malek,J., Fujii,C.,					
	Gerrard,C., Leech,V., de Jong,P., Ullu,E., Melville,S., Donelson.J.,					
	Fraser.C. and Adams.M.					
TITLE	Determination of clone end sequences from <i>Trypanosoma brucei</i> GUTat					
JOURNAL	10.1 sheared DNA library					
COMMENT	Unpublished (1999)					
	Other_GSSs: Sheared DNA-34021.TF					
	Contact: Najib M. El-Sayed					
	Department of Eukaryotic Genomics					
	The Institute for Genomic Research					
	9712 Medical Center Dr., Rockville, MD 20850, USA					
	Tel.: 301 838 0200					
	Fax: 301 838 0208					
	Email: nelsayed@tifgr.org					
	Clones are derived from the <i>Trypanosoma brucei</i> GUTat 10.1 sheared					
	DNA library constructed at TIGR. Clones will be available for					
	distribution through ATCC. Sheared DNA end sequences search page:					
	http://www.tigr.org/tldb/mdb/tbdb/ .					
	Seq primer: M13-Reverse					
	Class: shotgun.					
FEATURES	location/Qualifiers					
SOURCE	1..640					
	/organism="Trypanosoma brucei"					
	/strain="TREU927/4 GUTat 10.1"					
	/db_xref="taxon:5691"					
	/clone="Sheared DNA-34021"					
	/clone.lib="Sheared DNA"					
	/note="vector: pUC18; Site.1: SmaI: Constructed at The					
	Institute for Genomic Research (TIGR), Rockville, MD.					
	Genomic DNA isolated from a cloned population of					
	<i>Trypanosoma brucei</i> (TREU927/4 GUTat 10.1) was mechanically					
	sheared to give a tight size distribution (approx 2 kb).					
	The v + i method used for the library construction is					
	described in detail in Smith, H.O. and Venter, J.C.					
	(Making small insert libraries for whole genome shotgun					
	sequencing projects. In Genome Sequencing: A Practical					
	Approach, eds. M. Vaubin and B. Barrell, Oxford University					
	Press, 1999)."					
BASE COUNT	150 a 168 c 190 g 132 t					
ORIGIN						
Query Match	31.5%, Score 31.8; DB 17; Length 640;					
Best Local Similarity	64.0%; Pred. No. 90;					
Matches	48; Conservative 0; Mismatches 27; Indels 0; Gaps 0;					
OY	26 TATTTCGACCGCTGTACGCCGAGAGGGCGTGCTGCAGCAGAGGGATCGCC 85					
Dd	199 TATTCTGTCCGTGTACGCCGAGCGCGTAGCTGCAGCAGCATGTGAACCTT 258					
OY	86 ACGTAGTGACTCGC 100					
Dd	259 GCGTGGAGGATATGCT 273					
RESULT 15						
BH839856/c	737 bp DNA linear GSS 28-MAY-2002					
LOCUS	BH839856 LMKR200005F01f Zea mays L. Zea mays genomic clone LMKR200005F01f,					
DEFINITION	DNA sequence.					
ACCESSION	BH839856					
VERSION	BH839856.1 GI:21237924					
KEYWORDS	GSS.					
SOURCE	Zea mays.					
ORGANISM	Zea mays					

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 737)
Kim, S.W., Yu, Y., Lee, M.C., Main, D. and Wing, R.A.
Methyl-filtration genomic sequence from maize
Unpublished (2002)
Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: twing@clemson.edu
Total High Quality bases = 402
Seq primer: TAAATAGACTACTATAGGG
Class: shotgun
High quality sequence start: 10
High quality sequence stop: 520.
Location/Qualifiers

FEATURES
source

1. 737
/organism="Zea mays"
/strain="B73"
/db_xref="taxon:4577"
/clone="LMCR200005F01f"
/clone_lib="Zea mays L."
/issue_type="Leaf"
/lab_host="DH10B"
/note="Vector: pGEM-T easy; Site 1: Mcr BC;
Methyl-filtration library; Nuclei DNA was completely
digested with Mcr BC, size fractionated and transformed
to E.Coli.DH10B."

BASE COUNT 117 a 249 c 210 g 158 t 3 others
ORIGIN

Query Match 31.5%; Score 31.8; DB 17; Length 737;
Best Local Similarity 61.4%; Pred. No. 92;
Matches 51; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 19 GCGAGATATTTCGACCGCTGTGACCGCGAGAGCGGCGTGGCTGCGGAGAGGG 78
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 723 GCGAGAGTTTACGCGACCGCGAAGCCGCGAGAGCGGTAGTGGAAGGGAAGAGT 664
QY 79 ATGCGCACGCTGATGACTGCG 101
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 663 GTAGAGGCGGTTGGCTTGGC 641

Search completed: November 6, 2002, 15:52:44
Job time : 1201.62 secs


```
Query Match      32.7%; Score 33; DB 10; Length 594;
Best Local Similarity 63.0%; Pred. No. 0.18;
Matches 51; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

QY 15 GCGCGCGGAGATTTTCCAGCCCTGTGACGCCGAGAGAACGGCGCTGGCGCGACGA 74
DB 222 CAGCCCGGAGTCTCGCGCCAGCGCCGAGAGGCGTGGCGCTTGGCGCTT 163

QY 75 GGGATCGCGCAGCATGATGA 95
DB 162 GAGGATGGCTTCTCGACGCA 142

RESULT 2
US-09-832-320-3
; Sequence 3, Application US/09832320
; Patent No. US20010049834A1
; GENERAL INFORMATION:
; APPLICANT: Crane, Edmund H.
; APPLICANT: Crane, Virginia C.
; TITLE OF INVENTION: Maize Pathogenesis-Related
; FILE REFERENCE: 35718/214291
; CURRENT APPLICATION NUMBER: US/09/832,320
; PRIOR FILING DATE: 2001-04-10
; PRIOR APPLICATION NUMBER: US 60/195,801
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 612
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(612)
US-09-832-320-3

Query Match      28.9%; Score 29.2; DB 10; Length 612;
Best Local Similarity 65.2%; Pred. No. 2.3;
Matches 43; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 31 GCGACCGCTGTACGCCGAGAGCGGCGTGGCTGCGAGCAGGAGATCGCGACGCTG 90
DB 283 GCGCGCGCGGACGCTGGCGGACGAGCGCGCAGGCGCGGCGGTTCGCGGAGCGTG 342

QY 91 ATGGAC 96
DB 343 GGGGCC 348

RESULT 3
US-09-832-320-1
; Sequence 1, Application US/09832320
; Patent No. US20010049834A1
; GENERAL INFORMATION:
; APPLICANT: Crane, Edmund H.
; APPLICANT: Crane, Virginia C.
; TITLE OF INVENTION: Maize Pathogenesis-Related
; FILE REFERENCE: 35718/214291
; CURRENT APPLICATION NUMBER: US/09/832,320
; PRIOR FILING DATE: 2001-04-10
; PRIOR APPLICATION NUMBER: US 60/195,801
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 898
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
```

```
; NAME/KEY: CDS
; LOCATION: (63)...(674)
US-09-832-320-1

Query Match      28.9%; Score 29.2; DB 10; Length 898;
Best Local Similarity 65.2%; Pred. No. 2.3;
Matches 43; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 31 GCGACCGCTGTGTACGCCGAGAGCGGCGTGGCTGCGAGCAGGAGATCGCGACGCTG 90
DB 345 GCGCGCGCGGAGACGGGTGGCGGACGAGCGCGGCGGCGGCGGTTCGCGGAGCGTG 404

QY 91 ATGGAC 96
DB 405 GGGGCC 410

RESULT 4
US-09-741-669-159
; Sequence 159, Application US/09741669
; Patent No. US20020022718A1
; GENERAL INFORMATION:
; APPLICANT: Forsyth, R. Allyn
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; TITLE OF INVENTION: Genes identified as required for
; FILE REFERENCE: ELITRA.009A
; CURRENT APPLICATION NUMBER: US/09/741,669
; PRIOR FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: US 60/173005
; PRIOR FILING DATE: 1999-12-23
; NUMBER OF SEQ ID NOS: 481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 159
; LENGTH: 1518
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1518)
US-09-741-669-159

Query Match      28.7%; Score 29; DB 10; Length 1518;
Best Local Similarity 71.7%; Pred. No. 2.7;
Matches 38; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 49 GAGACGGGCGTGTGCTGCGCGAGCAGGGGATGCGCAGCTGTGACTCGCG 101
DB 70 GAGAGCTGGCGAACCCTGCGCGAGCGGAGATTGCTTCCGGAACGATTTCG 122

RESULT 5
US-09-815-242-6188
; Sequence 6188, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
```

;; PRIOR FILING DATE: 2000-05-23
;; PRIOR APPLICATION NUMBER: 60/207,727
;; PRIOR FILING DATE: 2000-05-26
;; PRIOR APPLICATION NUMBER: 60/242,578
;; PRIOR FILING DATE: 2000-10-23
;; PRIOR APPLICATION NUMBER: 60/253,625
;; PRIOR FILING DATE: 2000-11-27
;; PRIOR APPLICATION NUMBER: 60/257,931
;; PRIOR FILING DATE: 2000-12-22
;; PRIOR APPLICATION NUMBER: 60/269,308
;; PRIOR FILING DATE: 2001-02-16
;; NUMBER OF SEQ ID NOS: 14110
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 6188
;; LENGTH: 1518
;; TYPE: DNA
;; ORGANISM: Escherichia coli
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: (1)...(1518)
US-09-815-242-6188

Query Match 28.7%; Score 29; DB 10; Length 1518;
Best Local Similarity 71.7%; Pred. No. 2.7;
Matches 38; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

OY 49 GAGAACGGCGGTGGCTGGCGAGAGGAGTCCGACGTGATGACTCGCG 101
DB 70 GAGAGCTGGCGAACCTGGCGAGGAGGAGTGGCTTCGCGAACGATTTCCG 122

RESULT 6
US-09-815-242-7886

;; Sequence 7886, Application US/09815242
;; Patent No. US20020061569A1
;; GENERAL INFORMATION:
;; APPLICANT: Haselbeck, Robert
;; APPLICANT: Ohlson, Karl L.
;; APPLICANT: Zyskind, Judith W.
;; APPLICANT: Wall, Daniel
;; APPLICANT: Trawick, John D.
;; APPLICANT: Carr, Grant J.
;; APPLICANT: Yamamoto, Robert T.
;; APPLICANT: Xu, H. Howard
;; TITLE OF INVENTION: Identification of Essential Genes in
;; FILE REFERENCE: ELITRA.011A
;; CURRENT APPLICATION NUMBER: US/09/815,242
;; PRIOR FILING DATE: 2001-03-21
;; PRIOR APPLICATION NUMBER: 60/191,078
;; PRIOR FILING DATE: 2000-03-21
;; PRIOR APPLICATION NUMBER: 60/206,848
;; PRIOR FILING DATE: 2000-05-23
;; PRIOR APPLICATION NUMBER: 60/207,727
;; PRIOR FILING DATE: 2000-05-26
;; PRIOR APPLICATION NUMBER: 60/242,578
;; PRIOR FILING DATE: 2000-10-23
;; PRIOR APPLICATION NUMBER: 60/253,625
;; PRIOR FILING DATE: 2000-11-27
;; PRIOR APPLICATION NUMBER: 60/257,931
;; PRIOR FILING DATE: 2000-12-22
;; PRIOR APPLICATION NUMBER: 60/269,308
;; PRIOR FILING DATE: 2001-02-16
;; NUMBER OF SEQ ID NOS: 14110
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 7886
;; LENGTH: 1056
;; TYPE: DNA
;; ORGANISM: Pseudomonas aeruginosa
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: (1)...(1056)
US-09-815-242-7886

Query Match 27.1%; Score 27.4; DB 10; Length 1056;
Best Local Similarity 57.6%; Pred. No. 7.6;
Matches 49; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

OY 1 ATCGCCCGCCACCTCGCGCGAGATATTGGACCGCTGTACCGCGGAGAGCGGCG 60
DB 718 ACCGCGTTGCTCTGCTGTCGCGGAGTCCCGCAACCGCGCTACCGAGCGCTGAAG 777
OY 61 TGGCTGCGCGAGCAGGAGTGGCGC 85
DB 778 GGGCGCGCGCGCGCGCGCACCGCGC 802

RESULT 7
US-09-764-847-1301

;; Sequence 1301, Application US/09764847
;; Patent No. US20020132767A1
;; GENERAL INFORMATION:
;; APPLICANT: Rosen et al.
;; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;; FILE REFERENCE: PC009
;; CURRENT APPLICATION NUMBER: US/09/764,847
;; PRIOR FILING DATE: 2001-01-17
;; PRIOR APPLICATION data removed - consult PALM or file wrapper
;; NUMBER OF SEQ ID NOS: 2003
;; SOFTWARE: PatentIn Ver. 2.0
;; SEQ ID NO 1301
;; LENGTH: 15987
;; TYPE: DNA
;; ORGANISM: Homo sapiens
US-09-764-847-1301

Query Match 26.3%; Score 26.6; DB 10; Length 15987;
Best Local Similarity 56.2%; Pred. No. 14;
Matches 50; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

OY 7 CGCCACCTCGCGCGGAGATATTGCGACCGCTGTACCGCGGAGAGCGGCGGTG 66
DB 615 CGGCTCGCGGAGAGAGCGTCCGGCGCGGAGAGTGGTCTGGGGCTGGCGG 674
OY 67 CGCGAGCAGGAGATCGCGACGCTGATGGA 95
DB 675 GCGGAGTAGTAGGAGCGCGAGGCGGCTGGA 703

RESULT 8
US-09-960-352-1020/c

;; Sequence 1020, Application US/09960352
;; Patent No. US20020137139A1
;; GENERAL INFORMATION:
;; APPLICANT: Warren, Wesley C.
;; APPLICANT: Tao, Nengding
;; APPLICANT: Byatt, John C.
;; APPLICANT: Mathialagan, Nagappan
;; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
;; FILE REFERENCE: 16511.006/37-21(10298)C
;; CURRENT APPLICATION NUMBER: US/09/960,352
;; PRIOR FILING DATE: 2001-09-24
;; NUMBER OF SEQ ID NOS: 15112
;; SEQ ID NO 1020
;; LENGTH: 314
;; TYPE: DNA
;; ORGANISM: Bos taurus
;; OTHER INFORMATION: Clone ID: 05-LIB3058-001-Q1-K1-B1
US-09-960-352-1020

Query Match 26.1%; Score 26.4; DB 10; Length 314;
Best Local Similarity 57.1%; Pred. No. 14;
Matches 48; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

OY 5 CCGCGACCTCGCGCGGAGATATTGGAGCCGCTGTACCGCGGAGAGCGGCGGTGCG 64


```
Query Match          25.9%; Score 26.2; DB 10; Length 3773;
Best Local Similarity 56.3%; Pred. No. 17;
Matches 49; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY 8 GCCACCTCGCGCGAGATATTTCGACCGCTGTACCGCGAGAGCGCGCTGCTGC 67
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 952 GCCAGCGGACGAGAGAAAGTGGCCGCTGACACAGCTGGAGAAACCAAGACGCGGC 1011
QY 68 GCGAGCAGGAGATCGCGACGTGATGG 94
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1012 TGCAGCAGAGAGCTGTGACGACCTGCTGG 1038

RESULT 13
US-09-878-574-11248
; Sequence 11248, Application US/09878574
; Patent No. US20020110548A1
; GENERAL INFORMATION:
; APPLICANT: Byrum, Joseph R.
; APPLICANT: Thompson, Michael D.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(15401)B
; CURRENT APPLICATION NUMBER: US/09/878,574
; PRIOR FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 09/333,535
; NUMBER OF SEQ ID NOS: 15775
; SEQ ID NO 11248
; LENGTH: 259
; TYPE: DNA
; ORGANISM: Glycine max
; OTHER INFORMATION: Clone ID: 701064227H1
US-09-878-574-11248

Query Match          25.7%; Score 26; DB 10; Length 259;
Best Local Similarity 55.6%; Pred. No. 19;
Matches 50; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 6 CCGCCACCTCGCGCGAGATATTTCGACCGCTGTGACCGCGAGAGCGCGCTGGCT 65
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 50 CCAGGACCTCGCTGCGCGGATGTAACAGCGCGGCTGGGAGAGAGCGCGCAT 109
QY 66 GCGGAGCAGGAGATCGCGACGTGATGA 95
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 110 GGAGAGCTCTGATCGACGACATGCTGCA 139

RESULT 14
US-09-983-965-5707
; Sequence 5707, Application US/09983965
; Patent No. US20020137160A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 37-21(10297)C
; CURRENT APPLICATION NUMBER: US/09/983,965
; CURRENT FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: US 09/465,231
; PRIOR FILING DATE: 1999-12-15
; PRIOR APPLICATION NUMBER: US 60/113,678
; PRIOR FILING DATE: 1998-12-17
; NUMBER OF SEQ ID NOS: 5912
; SEQ ID NO 5707
; LENGTH: 316
; TYPE: DNA
; ORGANISM: Bos taurus
```

```
; FEATURE:
; OTHER INFORMATION: Clone ID: 56-LIB34-038-Q1-E1-F12
US-09-983-965-5707

Query Match          25.7%; Score 26; DB 10; Length 316;
Best Local Similarity 57.3%; Pred. No. 19;
Matches 47; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

QY 17 GCGCGAGATATTTCGACCGCTGTGACCGCGAGAGCGGCGCTGCGCGAGACAG 76
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 172 GCGCAGACGCTTTGTGTGACTGAACGTGAGCTGAGTGAGTGATGTGAGAGGTGT 231
QY 77 GGATCGCGACGTGATGACTC 98
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 232 CTGTCTACATGATGATGTCC 253

RESULT 15
US-09-822-830A-286/C
; Sequence 286, Application US/09822830A
; Patent No. US20020142952A1
; GENERAL INFORMATION:
; APPLICANT: Genetics Institute, Inc.
; APPLICANT: Wong, Gordon G.
; APPLICANT: Clark, Hilary
; APPLICANT: Fechtel, Kim
; APPLICANT: Agostino, Michael J.
; APPLICANT: Howes, Steven H.
; APPLICANT: Resnick, Richard J.
; APPLICANT: Gulukota, Kamalakara
; APPLICANT: Graham, James R.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
; FILE REFERENCE: GIN 6402
; CURRENT APPLICATION NUMBER: US/09/822,830A
; CURRENT FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: 60/195,604
; PRIOR FILING DATE: 2000-04-06
; NUMBER OF SEQ ID NOS: 631
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 286
; LENGTH: 1601
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc-feature
; LOCATION: 1490
; OTHER INFORMATION: n-a,c/g, or t
US-09-822-830A-286

Query Match          25.7%; Score 26; DB 10; Length 1601;
Best Local Similarity 58.7%; Pred. No. 19;
Matches 44; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 4 GCCGCCACCTCGCGCGAGATATTTCGACCGCTGTGACCGCGAGAGCGGCGCTGG 63
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1495 GCGTCNCCCCTGCGCGCGGAGATGCTGGACCGACGCTGGCGAGAGAGCGCAAGCTGC 1436
QY 64 CTGCGCAGCAGGAG 78
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1435 CCGCTGAGAGAGAGG 1421

Search completed: November 6, 2002, 20:32:52
Job time : 34.125 secs
```


GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 6, 2002, 11:51:54 : Search time 33.5 Seconds
(without alignments)
924.608 Million cell updates/sec

Title: US-09-724-876-2_COPY_36150_36250

Perfect score: 101

Sequence: 1 atgcgccgcacactcgcgcg.....gcgcacgtgatgacgcgcg 101

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : Listing first 45 summaries

- Issued_Patents_NA:*
- 1: /cgn2_6/ptodata/2/ina/5A.COMB.seq:*
 - 2: /cgn2_6/ptodata/2/ina/5B.COMB.seq:*
 - 3: /cgn2_6/ptodata/2/ina/6A.COMB.seq:*
 - 4: /cgn2_6/ptodata/2/ina/6B.COMB.seq:*
 - 5: /cgn2_6/ptodata/2/ina/PCrUS.COMB.seq:*
 - 6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	101	100.0	71989	US-09-443-501A-2	Sequence 2, Appl
2	99.4	98.4	68750	US-09-335-409-1	Sequence 1, Appl
3	99.4	98.4	68750	US-09-568-102-1	Sequence 1, Appl
4	99.4	98.4	68750	US-09-567-969-1	Sequence 1, Appl
5	99.4	98.4	68750	US-09-568-480-1	Sequence 1, Appl
6	99.4	98.4	68750	US-09-568-486-1	Sequence 1, Appl
7	99.4	98.4	68750	US-09-568-472-1	Sequence 1, Appl
8	99.4	98.4	68750	US-09-567-899-1	Sequence 1, Appl
9	43.4	43.0	4403765	US-09-103-840A-2	Sequence 2, Appl
10	43.4	43.0	4411529	US-09-103-840A-1	Sequence 1, Appl
11	32.4	32.1	28958	US-08-258-261B-6	Sequence 6, Appl
12	32.4	32.1	28958	US-08-456-837-6	Sequence 6, Appl
13	32.4	32.1	28958	US-08-457-342-6	Sequence 6, Appl
14	32.4	32.1	28958	US-08-457-464-6	Sequence 6, Appl
15	32.4	32.1	28958	US-08-458-076A-6	Sequence 6, Appl
16	32.4	32.1	28958	US-08-764-333A-4	Sequence 4, Appl
17	32.4	32.1	28958	US-08-457-335A-6	Sequence 6, Appl
18	32.4	32.1	28958	US-08-729-214-6	Sequence 6, Appl
19	32.4	32.1	28958	US-09-028-934-6	Sequence 6, Appl
20	32.4	32.1	49377	US-08-764-233A-1	Sequence 1, Appl
21	32	31.7	44377	US-08-804-227C-7	Sequence 7, Appl
22	32	31.7	44377	US-08-804-198-1	Sequence 1, Appl
23	29.6	29.3	884	US-09-219-391-11	Sequence 11, Appl
24	29.6	29.3	884	US-09-219-391-11	Sequence 11, Appl
25	29.6	29.3	1941	PCT-US91-07635-5	Sequence 5, Appl
26	29.6	29.3	6418	US-08-480-528A-11	Sequence 11, Appl
27	29.6	29.3	6418	US-08-479-666-11	Sequence 11, Appl

28	29.6	29.3	6418	PCT-US93-10520-11	Sequence 11, Appl
29	28.2	27.9	4257	US-08-690-473-1	Sequence 1, Appl
30	28.2	27.9	4257	US-09-259-821A-1	Sequence 1, Appl
31	28.2	27.9	4257	US-08-843-659-1	Sequence 1, Appl
32	28.2	27.9	12001	US-08-458-568A-11	Sequence 11, Appl
33	28.2	27.9	4411529	US-09-103-840A-1	Sequence 1, Appl
34	28	27.7	967	US-09-248-135-51	Sequence 51, Appl
35	27.8	27.5	43280	US-08-804-227C-1	Sequence 1, Appl
36	27.6	27.3	1914	US-09-291-922-19	Sequence 19, Appl
37	27.6	27.3	13987	US-08-804-227C-13	Sequence 13, Appl
38	27.6	27.3	44377	US-08-804-198-1	Sequence 1, Appl
39	27.6	27.3	44377	US-08-804-198-1	Sequence 1, Appl
40	27.2	26.9	2751	US-09-037-190-45	Sequence 45, Appl
41	27.2	26.9	2751	US-09-037-192-45	Sequence 45, Appl
42	27.2	26.9	2751	US-09-037-143-45	Sequence 45, Appl
43	27.2	26.9	2751	US-09-049-691-45	Sequence 45, Appl
44	27.2	26.9	2751	US-08-260-174-45	Sequence 45, Appl
45	27.2	26.9	2751	US-09-338-128A-45	Sequence 45, Appl

ALIGNMENTS

RESULT 1
US-09-443-501A-2
Sequence 2, Application US/09443501A
Patent No. 6303342
GENERAL INFORMATION:
APPLICANT: Kusan Biosciences, Inc.
APPLICANT: Julien, Bryan
APPLICANT: Khosla, Chaitan
APPLICANT: Tang, Li
APPLICANT: Ziermann, Rainer
TITLE OF INVENTION: Recombinant Methods and Materials for Producing
FILE REFERENCE: 30062-20031.00
CURRENT APPLICATION NUMBER: US/09/443,501A
CURRENT FILING DATE: 1999-11-19
PRIOR APPLICATION NUMBER: US 60/130,560
PRIOR FILING DATE: 1999-04-22
PRIOR APPLICATION NUMBER: US 60/122,620
PRIOR FILING DATE: 1999-03-03
PRIOR APPLICATION NUMBER: US 60/119,386
PRIOR FILING DATE: 1999-02-10
PRIOR APPLICATION NUMBER: US 60/109,401
PRIOR FILING DATE: 1998-11-20
NUMBER OF SEQ ID NOS: 22
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 71989
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Synthetic construct
US-09-443-501A-2
Query Match 100.0%; Score 101; DB 4; Length 71989;
Best Local Similarity 100.0%; Pred. No. 1;le-19;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGCGCCGCCACACCTCGCGCGGAGATATTTCGACCCGTGTGACCGCGGAGAAGCGCG 60
DB 36150 ATGCGCCGCCACACCTCGCGCGGAGATATTTCGACCCGTGTGACCGCGGAGAAGCGCG 36209
QY 61 TGGCTGCGCGAGCAGAGGGATGCGACGTGATGACACTCGCG 101
DB 36210 TGGCTGCGCGAGCAGAGGGATGCGACGTGATGACACTCGCG 36250
RESULT 2
US-09-335-409-1
Sequence 1, Application US/09335409

```
; Patent No. 6121029
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goelbach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/335,409
; PRIORITY FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-335-409-1
```

```
Query Match          98.4%; Score 99.4; DB 4; Length 68750;
Best Local Similarity 99.0%; Pred. No. 3e-19;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 ATGCCCCGACCTCGCGCGGAGATATTGGACCGCTGTGACCGCGGAGAGCGGCG 60
DB 41762 ATGCCCCGACCTCGCGCGGAGATATTGGACCGCTGTGACCGCGGAGAGCGGCG 41821
QY 61 TGGCTGCGGAGCAGGAGATCGCGCACGTGATGACTCGCG 101
DB 41822 TGGCTGCGGAGCAGGAGATCGCGCACGTGATGACTCGCG 41862
```

```
RESULT 3
US-09-568-102-1
; Sequence 1, Application US/09568102
; Patent No. 6346404
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goelbach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,102
; PRIORITY FILING DATE: 2000-05-10
; CURRENT APPLICATION NUMBER: 09/335,409
; PRIORITY FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-102-1
```

```
Query Match          98.4%; Score 99.4; DB 4; Length 68750;
Best Local Similarity 99.0%; Pred. No. 3e-19;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 ATGCCCCGACCTCGCGCGGAGATATTGGACCGCTGTGACCGCGGAGAGCGGCG 60
DB 41762 ATGCCCCGACCTCGCGCGGAGATATTGGACCGCTGTGACCGCGGAGAGCGGCG 41821
QY 61 TGGCTGCGGAGCAGGAGATCGCGCACGTGATGACTCGCG 101
DB 41822 TGGCTGCGGAGCAGGAGATCGCGCACGTGATGACTCGCG 41862
```

```
RESULT 4
US-09-567-969-1
```

```
; Sequence 1, Application US/09567969
; Patent No. 6355457
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goelbach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/567,969
; PRIORITY FILING DATE: 2000-05-10
; CURRENT APPLICATION NUMBER: 09/335,409
; PRIORITY FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-567-969-1
```

```
Query Match          98.4%; Score 99.4; DB 4; Length 68750;
Best Local Similarity 99.0%; Pred. No. 3e-19;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 ATGCCCCGACCTCGCGCGGAGATATTGGACCGCTGTGACCGCGGAGAGCGGCG 60
DB 41762 ATGCCCCGACCTCGCGCGGAGATATTGGACCGCTGTGACCGCGGAGAGCGGCG 41821
QY 61 TGGCTGCGGAGCAGGAGATCGCGCACGTGATGACTCGCG 101
DB 41822 TGGCTGCGGAGCAGGAGATCGCGCACGTGATGACTCGCG 41862
```

```
RESULT 5
US-09-568-480-1
; Sequence 1, Application US/09568480
; Patent No. 6355458
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goelbach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,480
; PRIORITY FILING DATE: 2000-05-10
; CURRENT APPLICATION NUMBER: 09/335,409
; PRIORITY FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-480-1
```

```
Query Match          98.4%; Score 99.4; DB 4; Length 68750;
Best Local Similarity 99.0%; Pred. No. 3e-19;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 ATGCCCCGACCTCGCGCGGAGATATTGGACCGCTGTGACCGCGGAGAGCGGCG 60
DB 41762 ATGCCCCGACCTCGCGCGGAGATATTGGACCGCTGTGACCGCGGAGAGCGGCG 41821
QY 61 TGGCTGCGGAGCAGGAGATCGCGCACGTGATGACTCGCG 101
DB 41822 TGGCTGCGGAGCAGGAGATCGCGCACGTGATGACTCGCG 41862
```

```
RESULT 6
US-09-568-486-1
; Sequence 1, Application US/09568486
; Patent No. 635459
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHLONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,486
; PRIOR FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-486-1

Query Match          98.4%; Score 99.4; DB 4; Length 68750;
Best Local Similarity 99.0%; Pred. No. 3e-19;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATCGCCCGCCACCTGCGGCGGAGATATTGCGACCGCTGTGACCGCGAGAGCGGCG 60
DB 41762 ATCGCCCGCCACCTGCGGCGGAGATATTGCGACCGCTGTGACCGCGAGAGCGGCG 41821
QY 61 TGGCTGCGGAGACGAGGATGCGCAGCTGATGACTCGCG 101
DB 41822 TGGCTGCGGAGACGAGGATGCGCAGCTGATGACTCGCG 41862

RESULT 7
US-09-568-472-1
; Sequence 1, Application US/09568472
; Patent No. 6358719
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHLONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,472
; PRIOR FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-472-1

Query Match          98.4%; Score 99.4; DB 4; Length 68750;
Best Local Similarity 99.0%; Pred. No. 3e-19;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATCGCCCGCCACCTGCGGCGGAGATATTGCGACCGCTGTGACCGCGAGAGCGGCG 60
DB 41762 ATCGCCCGCCACCTGCGGCGGAGATATTGCGACCGCTGTGACCGCGAGAGCGGCG 41821
QY 61 TGGCTGCGGAGACGAGGATGCGCAGCTGATGACTCGCG 101
DB 41822 TGGCTGCGGAGACGAGGATGCGCAGCTGATGACTCGCG 41862
```

```
DB 41822 TGGCTGCGGAGACGAGGATGCGCAGCTGATGACTCGCG 41862

RESULT 8
US-09-567-899-1
; Sequence 1, Application US/09567899
; Patent No. 6383787
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHLONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/567,899
; PRIOR FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-567-899-1

Query Match          98.4%; Score 99.4; DB 4; Length 68750;
Best Local Similarity 99.0%; Pred. No. 3e-19;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATCGCCCGCCACCTGCGGCGGAGATATTGCGACCGCTGTGACCGCGAGAGCGGCG 60
DB 41762 ATCGCCCGCCACCTGCGGCGGAGATATTGCGACCGCTGTGACCGCGAGAGCGGCG 41821
QY 61 TGGCTGCGGAGACGAGGATGCGCAGCTGATGACTCGCG 101
DB 41822 TGGCTGCGGAGACGAGGATGCGCAGCTGATGACTCGCG 41862

RESULT 9
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; PRIOR FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match          43.0%; Score 43.4; DB 4; Length 4403765;
Best Local Similarity 64.4%; Pred. No. 0.00079;
Matches 65; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 1 ATCGCCCGCCACCTGCGGCGGAGATATTGCGACCGCTGTGACCGCGAGAGCGGCG 60
```

Db 4287107 ATAGCCGGCGCGCGCCCTAGATCTTCGCCACCGCGGTACCCACAGCGTGCAGAG 4287048
QY 61 TGGCTGGCGAGCAGGGGATCGCGACGCTGATGACTCGCG 101
Db 4287047 CTAATGCCACATGGGTAATTGAGCATGTCTATGACTCGCG 4287007

RESULT 10

US-09-103-840A-1/C
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match 43.0%; Score 43.4; DB 4; Length 4411529;
Best Local Similarity 64.4%; Pred. No. 0.00079;
Matches 65; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 1 ATGCCCCGCACTCGCGCGGAGATTTGCGACCGCTGTACGCCGGAAGCGGCG 60
Db 4294850 ATAGCCGGCGCGCGCCCTAGATCTTCGCCACCGCGGTACCCACAGCGTGCAGAG 4294791
QY 61 TGGCTGGCGAGCAGGGGATCGCGACGCTGATGACTCGCG 101
Db 4294790 CTAATGCCACATGGGTAATTGAGCATGTCTATGACTCGCG 4294750

RESULT 11

US-08-258-261B-6
; Sequence 6, Application US/08258261B
; Patent No. 5639949
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James M.
; APPLICANT: Beck, James Joseph
; APPLICANT: HILL, Dwight Steven
; APPLICANT: RYALS, John Andrew
; APPLICANT: Gaffney, Thomas Deane
; APPLICANT: Lam, Stephen Ting
; APPLICANT: Hammer, Phillip E.
; APPLICANT: Uknes, Scott Joseph
; TITLE OF INVENTION: Genes for the synthesis of
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/258,261B
; FILING DATE: 08-JUN-1994
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/457,205
; FILING DATE: 01-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8614
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28958 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
US-08-258-261B-6

Query Match 32.1%; Score 32.4; DB 1; Length 28958;
Best Local Similarity 72.4%; Pred. No. 0.72;
Matches 42; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 2 TCGCCCCGCACTCGCGCGGAGATTTGCGACCGCTGTACGCCGGAAGCGGCG 59
Db 12950 TCGCAGCGCACCTCGGCGCGGAGTCTTCCACCGCCAGTCCAGGGAAGTGGAGCGC 13007

RESULT 12

US-08-456-837-6
; Sequence 6, Application US/08456837
; Patent No. 5643774
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James M.
; APPLICANT: Beck, James Joseph
; APPLICANT: HILL, Dwight Steven
; APPLICANT: RYALS, John Andrew
; APPLICANT: Gaffney, Thomas Deane
; APPLICANT: Lam, Stephen Ting
; APPLICANT: Hammer, Phillip E.
; APPLICANT: Uknes, Scott Joseph
; TITLE OF INVENTION: Genes for the synthesis of
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/456,837
; FILING DATE: 01-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/457,205
; FILING DATE: 01-JUN-1995
; APPLICATION NUMBER: 08/258,261
; FILING DATE: 08-JUN-1994
; ATTORNEY/AGENT INFORMATION:

```

: NAME: Elmer, James Scott
: REGISTRATION NUMBER: 36,129
: REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 919-541-8614
: TELEFAX: 919-541-8689
: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 28958 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: HYPOTHEetical: NO
: ANTI-SENSE: NO
US-08-456-837-6

```

```

Query Match 32.1%; Score 32.4; DB 1; Length 28958;
Best Local Similarity 72.4%; Pred. No. 0.72;
Matches 42; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

```

```

QY 2 TCGCCCGCCACCTCGGCGCGGAGATATTGCGACCGCTGGTACCGCGGAGAGCGGCG 59
Db 12950 TCGCAGCCACCTCGGCGCGGAGGCTTTGCGACCGCGGAGGAGTGAAGCGC 13007

```

```

RESULT 13
US-08-457-342-6
: Sequence 6, Application US/08457342
: Patent No. 5662898
: GENERAL INFORMATION:
: APPLICANT: Schupp, Thomas
: APPLICANT: Ligon, James M.
: APPLICANT: Beck, James Joseph
: APPLICANT: Hill, Dwight Steven
: APPLICANT: Ryals, John Andrew
: APPLICANT: Gaffney, Thomas Deane
: APPLICANT: Lam, Stephen Ting
: APPLICANT: Hammer, Phillip E.
: APPLICANT: Uknes, Scott Joseph
: TITLE OF INVENTION: Genes for the synthesis of
: TITLE OF INVENTION: antipathogenic substances
: NUMBER OF SEQUENCES: 22
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Ciba-Geigy Corporation
: STREET: 7 Skyline Drive
: CITY: Hawthorne
: STATE: NY
: COUNTRY: USA
: ZIP: 10532
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/457,342
: FILING DATE: 01-JUN-1995
: CLASSIFICATION: 424
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/457,205
: FILING DATE: 01-JUN-1995
: APPLICATION NUMBER: 08/258,261
: FILING DATE: 08-Jun-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Elmer, James Scott
: REGISTRATION NUMBER: 36,129
: REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 919-541-8614
: TELEFAX: 919-541-8689
: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:

```

```

: LENGTH: 28958 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: HYPOTHEtical: NO
: ANTI-SENSE: NO
US-08-457-342-6

```

```

Query Match 32.1%; Score 32.4; DB 1; Length 28958;
Best Local Similarity 72.4%; Pred. No. 0.72;
Matches 42; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

```

```

QY 2 TCGCCCGCCACCTCGGCGCGGAGATATTGCGACCGCTGGTACCGCGGAGAGCGGCG 59
Db 12950 TCGCAGCCACCTCGGCGCGGAGGCTTTGCGACCGCGGAGGAGTGAAGCGC 13007

```

```

RESULT 14
US-08-457-646A-6
: Sequence 6, Application US/08457646A
: Patent No. 5679560
: GENERAL INFORMATION:
: APPLICANT: Schupp, Thomas
: APPLICANT: Ligon, James M.
: APPLICANT: Beck, James Joseph
: APPLICANT: Hill, Dwight Steven
: APPLICANT: Ryals, John Andrew
: APPLICANT: Gaffney, Thomas Deane
: APPLICANT: Lam, Stephen Ting
: APPLICANT: Hammer, Phillip E.
: APPLICANT: Uknes, Scott Joseph
: TITLE OF INVENTION: Genes for the synthesis of
: TITLE OF INVENTION: antipathogenic substances
: NUMBER OF SEQUENCES: 22
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Ciba-Geigy Corporation
: STREET: 7 Skyline Drive
: CITY: Hawthorne
: STATE: NY
: COUNTRY: USA
: ZIP: 10532
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/457,646A
: FILING DATE: 01-JUN-1995
: CLASSIFICATION: 530
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/457,205
: FILING DATE: 01-JUN-1995
: APPLICATION NUMBER: 08/258,261
: FILING DATE: 08-Jun-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Elmer, James Scott
: REGISTRATION NUMBER: 36,129
: REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 919-541-8614
: TELEFAX: 919-541-8689
: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 28958 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: HYPOTHEtical: NO
: ANTI-SENSE: NO
US-08-457-646A-6

```

Query Match 32.1%; Score 32.4; DB 1; Length 28958;
Best Local Similarity 72.4%; Pred. No. 0.72;
Matches 42; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Search completed: November 6, 2002, 17:26:15
Job time: 1793.62 secs

Db 12950 TCGCAGCCACCTCGGCGCGAGTCTTTCACCGCCAGTCCAGGAGTGGAGCGC 13007

RESULT 15

US-08-458-076A-6
; Sequence 6, Application US/08458076A
; Patent No. 5698425
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Lyon, James M.
; APPLICANT: Beck, James Joseph
; APPLICANT: Hill, Dwight Steven
; APPLICANT: Ryals, John Andrew
; APPLICANT: Gaffney, Thomas Deane
; APPLICANT: Lam, Stephen Ting
; APPLICANT: Hammer, Phillip E.
; TITLE OF INVENTION: Genes for the synthesis of
; TITLE OF INVENTION: antipathogenic substances
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/458,076A
; FILING DATE: 01-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/457,205
; FILING DATE: 01-JUN-1995
; APPLICATION NUMBER: 08/258,261
; FILING DATE: 08-Jun-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8614
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28958 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; US-08-458-076A-6

Query Match 32.1%; Score 32.4; DB 1; Length 28958;
Best Local Similarity 72.4%; Pred. No. 0.72;
Matches 42; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Db 12950 TCGCAGCCACCTCGGCGCGAGTCTTTCACCGCCAGTCCAGGAGTGGAGCGC 13007

	C	17	31.4	31.1	7.50	24	ABQ37161	Oligonucleotide fold
	C	18	31.4	31.1	7.53	21	AAK14651	Nucleotide sequence
	C	19	30.6	30.3	6.54	22	AAF31637	Mycobacterium tuberc
	C	20	29.6	29.3	8.84	22	AAC89692	Human tissue morph
	C	21	29.6	29.3	1.94	13	AAO24522	cDNA encoding prepro
	C	22	29.6	29.3	6.418	15	AAO65395	Osteogenic protein
	C	23	29.6	29.3	7.7536	21	AAA14651	Nucleotide sequenc
	C	24	29.2	28.9	6.12	24	ABA96418	Maize PRL-C10 codi
	C	25	29.2	28.9	8.98	24	ABA96417	Maize PRL-C10 enc
	C	26	29	28.7	1.518	22	AAH81360	Escherichia coli F
	C	27	29	28.7	1.518	23	AAAS2551	E. coli DNA for cel
	C	28	29	28.7	1.214	22	AAAS46260	DNA encoding novel
	C	29	28.6	28.3	1.377	24	ABQ90092	M. capsulatus gene
	C	30	28.4	28.1	3.957	22	AAA09686	HSV-2 immediate ex
	C	31	28.4	28.1	1.54746	24	AAAO5916	Human herpesvirus
	C	32	28.4	28.1	1.54746	24	AAD25519	cDNA encoding maiz
	C	33	28.2	27.9	3.36	24	ABN21420	Human ORFX polynu
	C	34	28.2	27.9	1.684	23	AAAS90724	DNA encoding novel
	C	35	28.2	27.9	1.764	21	AAAS52589	cDNA encoding maiz
	C	36	28.2	27.9	1.780	21	AAAS52591	cDNA encoding maiz
	C	37	28.2	27.9	1.879	22	ABA83021	Human transcriptid
	C	38	28.2	27.9	1.884	22	ABA83033	Human transcriptid
	C	39	28.2	27.9	2.939	23	ABL10413	Drosophila melanog
	C	40	28.2	27.9	4.257	19	AAV68520	The nucleotide seq
	C	41	28.2	27.9	5.1427	19	AAV10362	Infected cell prot
	C	42	28.2	27.9	5.142	23	ABL10412	Drosophila melano
	C	44	28.2	27.9	6.750	22	AAK83688	Human immune/haema
	C	45	28.2	27.9	1.2001	16	AAQ76213	HSV L/SR region.

ALIGNMENTS

RESULT_1	
ID	AAA29349 standard; DNA; 71989 BP.
XX	AAA29349;
XX	
DT	12-SEP-2000 (first entry)
XX	
DE	Sorangium cellulorum epothilone polyketide synthase operon genomic DNA.
XX	
KW	epothilone; polyketide synthase; ePOA; ePOB; ePOC; ePOD; ePOE; ePOF; Km Epox.; epOK; P450 epoxidase; ORFA; ORPB; promoter; enhancer; anti-fungal; tubulin polymerization assay; anti-tumour; cytostatic; ds.
KM	
KX	
OS	Sorangium cellulorum.
XX	
FH	Location/Qualifiers
FT	3..992
FT	/tag= a
FT	/label= ORF_A
FT	/product= transposase
FT	/note= "not part of the PKS"
FT	989..1501
FT	/tag= b
FT	/label= ORF_B
FT	/product= transposase
FT	/note= "not part of the PKS"
FT	1998..6263
FT	/tag= c
FT	/label= epok_gene
FT	/note= "encodes the loading domain"
FT	2031..3548
misc_RNA	

```
FT      /tag= d
FT      /note= "encodes ketide synthase (KS-Y) of the loading
FT      domain"
FT      misc_RNA
FT      3621..4661
FT      /tag= e
FT      /note= "encodes acyl transferase (AT) of the loading
FT      domain"
FT      misc_RNA
FT      4917..5810
FT      /tag= f
FT      /note= "encodes enoyl reductase (ER) of the loading
FT      domain, potentially involved in formation of the
FT      thiazole moiety"
FT      misc_RNA
FT      5856..6135
FT      /tag= g
FT      /note= "encodes acyl carrier protein (ACP) of the loading
FT      domain"
FT      CDS
FT      6260..10493
FT      /tag= h
FT      /label= epob_gene
FT      /note= "encodes module 1, the NRPS module"
FT      misc_RNA
FT      2031..3548
FT      /tag= i
FT      /note= "encodes condensation domain C2 of the NRPS
FT      module"
FT      misc_RNA
FT      2031..3548
FT      /tag= j
FT      /note= "encodes condensation domain C2 of the NRPS
FT      module"
FT      misc_RNA
FT      6861..6887
FT      /tag= k
FT      /note= "encodes heterocyclization signature sequence"
FT      misc_RNA
FT      6861..6887
FT      /tag= l
FT      /note= "encodes condensation domain C4 of the NRPS
FT      module"
FT      misc_RNA
FT      7358..7366
FT      /tag= m
FT      /note= "encodes condensation domain C7 (partial) of the
FT      NRPS module"
FT      misc_RNA
FT      7898..7921
FT      /tag= n
FT      /note= "encodes adenylation domain A1 of the NRPS module"
FT      misc_RNA
FT      7898..7921
FT      /tag= o
FT      /note= "encodes adenylation domain A1 of the NRPS module"
FT      misc_RNA
FT      8261..8308
FT      /tag= p
FT      /note= "encodes adenylation domain A3 of the NRPS module"
FT      misc_RNA
FT      8411..8422
FT      /tag= q
FT      /note= "encodes adenylation domain A4 of the NRPS module"
FT      misc_RNA
FT      8861..8905
FT      /tag= r
FT      /note= "encodes adenylation domain A6 of the NRPS module"
FT      misc_RNA
FT      8966..8983
FT      /tag= s
FT      /note= "encodes adenylation domain A7 of the NRPS module"
FT      misc_RNA
FT      9090..9179
FT      /tag= t
FT      /note= "encodes adenylation domain A8 of the NRPS module"
FT      misc_RNA
FT      9183..9992
FT      /tag= u
FT      /note= "encodes oxidation region for forming thiazole"
FT      misc_RNA
FT      10121..10138
FT      /tag= v
FT      /note= "encodes adenylation domain A10 of the NRPS
FT      module"
FT      misc_RNA
FT      10261..10306
FT      /tag= w
FT      /note= "encodes thiolation domain (PCP) of the NRPS
FT      module"
FT      CDS
FT      10639..16137
FT      /tag= x

FT      /label= epoc_gene
FT      /note= "encodes module 2"
FT      misc_RNA
FT      10654..12033
FT      /tag= y
FT      /note= "encodes KS2, the KS domain of module 2"
FT      misc_RNA
FT      12250..13287
FT      /tag= z
FT      /note= "encodes AT2, the AT domain of module 2"
FT      misc_RNA
FT      13327..13899
FT      /tag= aa
FT      /note= "encodes dehydratase (DH) 2, the DH domain of
FT      module 2"
FT      misc_RNA
FT      14962..15756
FT      /tag= ab
FT      /note= "encodes ketoreductase (KR) 2, the KR domain of
FT      module 2"
FT      misc_RNA
FT      15763..16008
FT      /tag= ac
FT      /note= "encodes ACP2, the ACP domain of module 2"
FT      CDS
FT      16134..37907
FT      /tag= ad
FT      /label= epob_gene
FT      /note= "encodes modules 3-6"
FT      misc_RNA
FT      16425..17606
FT      /tag= ae
FT      /note= "encodes KS3"
FT      misc_RNA
FT      17817..18857
FT      /tag= af
FT      /note= "encodes AT3"
FT      misc_RNA
FT      19581..20396
FT      /tag= ag
FT      /note= "encodes KR3"
FT      misc_RNA
FT      20424..20642
FT      /tag= ah
FT      /note= "encodes ACP3"
FT      misc_RNA
FT      20706..22082
FT      /tag= ai
FT      /note= "encodes KS4"
FT      misc_RNA
FT      22296..23336
FT      /tag= aj
FT      /note= "encodes AT4"
FT      misc_RNA
FT      24069..24647
FT      /tag= ak
FT      /note= "encodes KR4"
FT      misc_RNA
FT      24867..25151
FT      /tag= al
FT      /note= "encodes ACP4"
FT      misc_RNA
FT      25203..26576
FT      /tag= am
FT      /note= "encodes KS5"
FT      misc_RNA
FT      26793..27883
FT      /tag= an
FT      /note= "encodes AT5"
FT      misc_RNA
FT      27966..28574
FT      /tag= ao
FT      /note= "encodes DHS"
FT      misc_RNA
FT      29433..30287
FT      /tag= ap
FT      /note= "encodes ER5"
FT      misc_RNA
FT      30321..30869
FT      /tag= aq
FT      /note= "encodes KR5"
FT      misc_RNA
FT      31077..31373
FT      /tag= ar
FT      /note= "encodes ACP5"
FT      misc_RNA
FT      31440..32807
FT      /tag= as
FT      /note= "encodes KS6"
FT      misc_RNA
FT      33018..34067
FT      /tag= at
FT      /note= "encodes AT6"
FT      misc_RNA
FT      34107..34676
FT      /tag= au
```

```
FT /note= "encodes DH6"
FT misc_RNA
FT 35760..36641
FT /tag= av
FT /note= "encodes ER6"
FT misc_RNA
FT 36705..37256
FT /tag= aw
FT /note= "encodes KR6"
FT misc_RNA
FT 37470..37769
FT /tag= ax
FT /note= "encodes ACP6"
FT CDS
FT 37912..49308
FT /tag= ay
FT /label= "epoE_gene"
FT /note= "encodes modules 7 and 8"
FT misc_RNA
FT 38014..39375
FT /tag= az
FT /note= "encodes KS7"
FT misc_RNA
FT 39589..40626
FT /tag= ba
FT /note= "encodes AT7"
FT misc_RNA
FT 41341..41922
FT /tag= db
FT /note= "encodes KR7"
FT misc_RNA
FT 42181..42423

Query Match      100.0%; Score 101; DB 21; Length 71989;
Best Local Similarity 100.0%; Pred. No. 8,4e-19;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCGCCGCGACCTGCGCGGAGATATTGGACCGCTGTGTACCGCGAGAGAGCGGCG 60
Db 36150 ATGCGCCGCGACCTGCGCGGAGATATTGGACCGCTGTGTACCGCGAGAGAGCGGCG 36209
QY 61 TGGCTGCCGCGAGCAGCGGATCGCGACGTGATGACTCGCG 101
Db 36210 TGGCTGCCGCGAGCAGCGGATCGCGACGTGATGACTCGCG 36250

RESULT 2
AA255887
ID AA255887 standard; DNA; 68750 BP.
XX
AC AA255887;
XX
DT 10-APR-2000 (first entry)
XX
DE Sorangium cellulosum 68.75 kb contig.
XX
KW Epothilone biosynthesis; type I polyketide synthase; taxol substitute; anticancer; ds.
XX
OS Sorangium cellulosum.
XX
FH Location/Qualifiers
FT 1..1826
FT CDS
FT /tag= a
FT /partial
FT /product= "Partial Orf 1 protein (AAV58580)"
FT /note= "No initiation codon given in the specification"
FT CDS
FT /tag= b
FT /product= "Orf 2 protein (AAV58581)"
FT 3415..5556
FT /tag= c
FT /product= "Orf 3 protein (AAV58582)"
FT CDS
FT /tag= d
FT /product= "Orf 4 protein (AAV58583)"
FT 6226..6675
FT /tag= e
FT /product= "Orf 5 protein (AAV58584)"
FT 7610..11875
FT /tag= f
```

```
FT /product= "Type I polyketide synthase, EPOS A"
FT 11872..116104
FT /tag= g
FT /product= "Non-ribosomal peptide synthetase, EPOS P"
FT 16251..21749
FT /tag= h
FT /product= "Type I polyketide synthase, EPOS B"
FT 21746..43519
FT /tag= i
FT /product= "Type I polyketide synthase, EPOS C"
FT 43524..54920
FT /tag= j
FT /product= "Type I polyketide synthase, EPOS D"
FT 54935..62254
FT /tag= k
FT /product= "Type I polyketide synthase, EPOS E"
FT 62369..63628
FT /tag= l
FT /product= "Cytochrome P450 oxygenase homologue, EPOS F"
FT 63779..64333
FT /tag= m
FT /product= "Orf 6 protein (AAV58585)"
FT CDS
FT /tag= n
FT /product= "Orf 7 protein (AAV58586)"
FT 64363..64920
FT /tag= o
FT /product= "Orf 8 protein (AAV58587)"
FT CDS
FT /tag= p
FT /product= "Orf 9 protein (AAV58588)"
FT 65063..65767
FT /tag= q
FT /product= "Orf 10 protein"
FT CDS
FT /tag= r
FT /product= "Orf 11 protein (AAV58590)"
FT CDS
FT /tag= s
FT /product= "Orf 12 protein (AAV58591)"
FT 66667..67137
FT /tag= t
FT /product= "Orf 13 protein (AAV58592)"
FT CDS
FT /tag= u
FT /product= "Orf 14 protein (AAV58593)"
FT CDS
FT /tag= v
FT /product= "Partial Orf 15 protein (AAV58594)"
FT /note= "No termination codon given in the specification"
FT W09966028-A2.
FT 23-DEC-1999.
FT PD
FT 16-JUN-1999; 99WO-EP04171.
FT PF
FT 18-JUN-1998; 98US-0099504.
FT PR
FT 24-SEP-1998; 98US-0101631.
FT PR
FT 05-FEB-1999; 99US-0118906.
FT XX
FT (NOVS ) NOVARTIS AG.
FT PA (NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.
FT XX Schupp T, Ligon JM, Molnar I, Zirkle R, Goerlach J, Cyr D;
```

XX	WPI: 2000-097741/08.	
DR	P-PSDB: AAY58573, AAY58574, AAY58575, AAY58576, AAY58577, AAY58578,	
DR	AAY58579, AAY58580, AAY58581, AAY58582, AAY58583, AAY58584,	
DR	AAY58585, AAY58586, AAY58587, AAY58588, AAY58590, AAY58591,	
XX	AAY58592, AAY58593, AAY58594.	
PT	New isolated epothilone synthase genes, used for the recombinant	
PT	production of epothilone for use in cancer therapy	
XX	Claim 14: Page 87-104; 174pp; English.	
XX	This sequence represents a 68.75 kb contig from <i>Sorangium cellulosum</i>	
CC	comprising 22 open reading frames (ORFs) and includes genes encoding	
CC	proteins involved in the biosynthesis of epothilones. Epothilones A and	
CC	B are 16-membered macrocyclic polyketides with an acylcysteine-derived	
CC	starter unit; polyketides being synthesised from two-carbon building	
CC	blocks, the beta-carbon of which always carries a keto group. Each round	
CC	of two-carbon addition is carried out by a complex of enzymes known as	
CC	the polyketide synthase in a manner similar to fatty acid biosynthesis.	
CC	EPOS A (AAY58573) and EPOS P (AAY58574) are involved in formation of	
CC	the thiazole ring formation of epothilones, and EPOS B, EPOS C, EPOS D	
CC	and EPOS E (AAY58575-Y58578) are involved in polyketide backbone	
CC	formation. EPO F (AAY58579) is an epothilone macro-lactone oxidase, and	
CC	the proteins Ort 3 (AAY58582) and Ort14 (AAY58593) are thought to be	
CC	involved in transport. Epothilones mimic the biological activity of	
CC	taxol, and may be substituted for taxol in cancer chemotherapeutic	
CC	compositions. Epothilones exhibit a much lower drop in potency against a	
CC	multiply drug-resistant cell line compared with taxol, and are	
CC	considerably less efficiently exported from such cells by the multidrug	
CC	resistance protein (MDR, or P-glycoprotein). Despite the potential of	
CC	epothilones as anticancer agents, they are problematical to produce on a	
CC	large scale. Epothilones are too complex for industrial scale chemical	
CC	synthesis, and <i>Sorangium cellulosum</i> is difficult to ferment, producing	
CC	poor yields of epothilones. The nucleic acids of the invention may be	
CC	used for the recombinant production of epothilones in a heterologous host	
CC	that is more amenable to fermentation.	
XX		
XX	Sequence 68750 BP: 9596 A; 22458 C; 25537 G; 11159 T; 0 other:	
SO		
	Query Match 98.4%; Score 99.4; DB 21; Length 68750;	
	Best Local Similarity 99.0%; Pred. No. 2, 3e-18;	
	Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0	
OY	1 ATCGCCCGCCACCTCGGCCGAGATATTTCGACCGCTGTCACCGCGGAGAACGGCGC 60	
DB	41762 ATTCGCCCGCCACCTCGGCCGAGATATTTCGACCGCTGTCACCGCGGAGAACGGCGG 41821	
OY	61 TGGCTGCGGAGACAGAGGATCGCGACGATGATGACTCGCG 101	
DB	41822 TGGCTGCGGAGACAGAGGATCGCGACGATGATGACTCGCG 41802	
RESULT 3		
AAAF90037	AAAF90037 standard; DNA; 6462 BP.	
XX	AAF90037:	
XX	06-AUG-2001 (first entry)	
DE	Nucleotide sequence of a type I polyketide synthase.	
XX	Metabolic pathway operon; polyketide; polyketide antibiotic;	
KW	type I polyketide synthase; ss.	
XX	unidentified.	
OS		
XX		
FH	Key Location/Qualifiers	
FT	CDS 1..6462	
FT	/*tag= a	
FT	/product= "type I polyketide synthase"	
FT	/transl_except= "(1..3, aa: Met)"	

PN	MO200140497-A2.
XX	
PD	07-JUN-2001.
XX	
PF	27-NOV-2000; 2000WO-FR03311.
XX	
PR	29-NOV-1999; 99FR-0015032.
XX	
PR	07-JUN-2000; 2000US-0209800.
XX	
PA	(AVET) AVENTIS PHARMA SA.
PI	Jeanm'n P, Pernodet J, Guerinneau M, Simonet P, Courtois S;
PI	Cappellano C, Franconu F, Raynal A, Ball M, Sezouev G, Tuphile K;
PI	Frostegard A;
DR	WPI: 2001-374849/39.
XX	
DR	P-PSDB; AAB83974.
XX	
PT	Collection of nucleic acids from environmental samples, useful for
PT	identifying e.-g. genes encoding polyketide synthases and derived
PT	antibiotics
XX	
PS	Claim 35; Page 307-309; 356pp; French.
XX	
CC	The specification describes a method for the preparation of a collection
CC	of nucleic acids from organisms in a soil sample. The method comprises
CC	milling a dried sample to produce microparticles; suspending these in
CC	liquid buffer; extraction of nucleic acids from the microparticle;
CC	passing nucleic acid-containing solution through a molecular sieve;
CC	passing nucleic acid-enriched fractions through an anion exchange
CC	chromatography material; and recovering fractions containing purified
CC	nucleic acids. The nucleic acids are sources for sequences that encode
CC	either operons involved in a metabolic pathway (specifically polyketide
CC	synthesis) or polypeptides, particularly for production of therapeutic
CC	or agricultural compounds, especially polyketide antibiotics. AAF90034-39
CC	represent open reading frames (ORFs) of the coding strand of cosmid
CC	a26g1, and encode type I polyketide synthases.
XX	
SO	Sequence 6462 BP; 1073 A; 2067 C; 2135 G; 1187 T; 0 other;
XX	
Query Match	52.5%; Score 53; DB 22; Length 6462;
Best Local Similarity	70.3%; Pred. No. 1,1e-05;
Matches	71; Conservative 0; Mismatches 30; Indels 0; Gaps 0;
OY	1 ATCGCCCGCACCTCGCGCGGAGATATTTCGACCGCTGTACGCCGAGAGCGGGCG 60
DB	4750 ATCGACAGCGCGCGGCGCGGAGATCTTCGTACTGCGCGGAGATCCGAAACGAGCG 4809
OY	61 TGGCTGCGGAGACGAGGATTCGCGGACGAGATGAGACTGCG 101
DB	4810 TATCTGCGCTCGCTGGGCAATCGCGCATGTTCGAGATCGCG 4850
RESULT 4	
AAF90033	
ID	AAF90033 standard; DNA; 34071 BP.
XX	
AC	AAF90033;
XX	
DT	06-AUG-2001 (first entry)
XX	
DE	Nucleotide sequence of cosmid a26g1 (coding strand).
XX	
KW	Metabolic pathway operon; polyketide; polyketide antibiotic; ss.
XX	
OS	Synthetic.
XX	
PN	MO200140497-A2.
XX	
PD	07-JUN-2001.
XX	
PF	27-NOV-2000; 2000WO-FR03311.

```
XX 29-NOV-1999; 99FR-0015032.
PR 07-JUN-2000; 2000US-0209800.
XX
XX (AVET ) AVENTIS PHARMA SA.
XX
PI Jeannin P, Pernodet J, Guerinneau M, Simonet P, Courtois S;
PI Cappellano C, Francou F, Raynal A, Ball M, Sezonov G, Tuphile K;
PI Frostegard A;
XX
XX WPI: 2001-374849/39.
XX
PT Collection of nucleic acids from environmental samples, useful for
PT identifying e.g. genes encoding polyketide synthases and derived
PT antibiotics
XX
PS Example 14; Page 289-300; 356pp; French.
XX
XX The specification describes a method for the preparation of a collection
XX of nucleic acids from organisms in a soil sample. The method comprises
XX milling a dried sample to produce microparticles; suspending these in
XX liquid buffer; extraction of nucleic acids from the microparticle;
XX passing nucleic acid-containing solution through a molecular sieve;
XX chromatography material; and recovering fractions containing purified
XX nucleic acids. The nucleic acids are sources for sequences that encode
XX either operons involved in a metabolic pathway (specifically polyketide
XX synthesis) or polypeptides, particularly for production of therapeutic
XX or agricultural compounds, especially polyketide antibiotics. The present
XX sequence represents cosmid a26g1 coding strand, which encodes different
XX polyketide synthases.
XX
SQ Sequence 34071 BP; 5791 A; 10858 C; 11089 G; 6333 T; 0 other;
XX
Query Match 52.5%; Score 53; DB 22; Length 34071;
Best Local Similarity 70.3%; Pred. No. 1.2e-05;
Matches 71; Conservative 0; Mismatches 30; Indels 0; Gaps 0;
XX
QY 1 ATCGCGCGCACCTGCGGCGGAGATATTGTGGACCGCTGGTACGCGGAGAGCGGGCG 60
Db 22973 ATCGCACAGCGTGGCGGCGGCGGAGATCTTCGCTACTGCGGGAGTCCGAAAAACGAGCG 23032
XX
QY 61 TGGCTGCGGAGACAGGGGATCGCGACGTGATGACTCGCG 101
Db 23033 TATCTGCGCTGCGTGGCGATCGCGATGTTCGATTCGCG 23073
XX
RESULT 5
AAF90032/C
ID AAF90032 standard; DNA; 42717 BP.
XX
XX AAF90032;
AC
XX
XX 06-AUG-2001 (first entry)
DT
XX
XX Nucleotide sequence of cosmid a26g1 (non-coding strand).
DE
XX
XX Metabolic pathway operon; polyketide; polyketide antibiotic; ss.
KM
XX
OS Synthetic.
XX
XX WO200140497-A2.
XX
XX 07-JUN-2001.
PD
XX
XX 27-NOV-2000; 2000WO-FR03311.
PF
XX
XX 29-NOV-1999; 99FR-0015032.
PR
XX
XX 07-JUN-2000; 2000US-0209800.
PR
XX
XX (AVET ) AVENTIS PHARMA SA.
PA
XX
XX Jeannin P, Pernodet J, Guerinneau M, Simonet P, Courtois S;
PI
```

```
PI Cappellano C, Francou F, Raynal A, Ball M, Sezonov G, Tuphile K;
PI Frostegard A;
XX
XX WPI: 2001-374849/39.
XX
PT Collection of nucleic acids from environmental samples, useful for
PT identifying e.g. genes encoding polyketide synthases and derived
PT antibiotics
XX
PS Example 14; Page 274-288; 356pp; French.
XX
XX The specification describes a method for the preparation of a collection
XX of nucleic acids from organisms in a soil sample. The method comprises
XX milling a dried sample to produce microparticles; suspending these in
XX liquid buffer; extraction of nucleic acids from the microparticle;
XX passing nucleic acid-containing solution through a molecular sieve;
XX chromatography material; and recovering fractions containing purified
XX nucleic acids. The nucleic acids are sources for sequences that encode
XX either operons involved in a metabolic pathway (specifically polyketide
XX synthesis) or polypeptides, particularly for production of therapeutic
XX or agricultural compounds, especially polyketide antibiotics. The present
XX sequence represents cosmid a26g1 (non-coding strand). The sense strand
XX encodes different polyketide synthases.
XX
SQ Sequence 42717 BP; 8230 A; 13520 C; 13184 G; 7782 T; 1 other;
XX
Query Match 52.5%; Score 53; DB 22; Length 42717;
Best Local Similarity 70.3%; Pred. No. 1.2e-05;
Matches 71; Conservative 0; Mismatches 30; Indels 0; Gaps 0;
XX
QY 1 ATCGCGCGCACCTGCGGCGGAGATATTGTGGACCGCTGGTACGCGGAGAGCGGGCG 60
Db 11147 ATCGCACAGCGTGGCGGCGGCGGAGATCTTCGCTACTGCGGGAGTCCGAAAAACGAGCG 11088
XX
QY 61 TGGCTGCGGAGACAGGGGATCGCGACGTGATGACTCGCG 101
Db 11087 TATCTGCGCTGCGTGGCGATCGCGCATGTTCGATTCGCG 11047
XX
RESULT 6
AAF31641
ID AAF31641 standard; DNA; 6333 BP.
XX
XX AAF31641;
AC
XX
XX 09-APR-2001 (first entry)
DT
XX
XX Mycobacterium tuberculosis mas gene.
DE
XX
XX Mycobacterium tuberculosis; attenuated microorganism;
KM
XX
XX signature tagged transposon mutant; mutant library;
KM
XX
XX vaccine; mas; mycoerotic acid synthase; ds.
KW
XX
XX Mycobacterium tuberculosis.
OS
XX
XX WO200102555-A1.
XX
XX 11-JAN-2001.
PD
XX
XX 06-JUL-2000; 2000WO-IB00950.
PF
XX
XX 06-JUL-1999; 99US-0142982.
PR
XX
XX 08-JUL-1999; 99US-0142833.
PR
XX
XX (INSP ) INST PASTEUR.
PA
XX
XX Gicquel B, Guilhot C, Camacho L;
PI
XX
XX WPI: 2001-091804/10.
DR
XX
XX P-PSDB; AAB66471.
DR
XX
```

PT Screening a mutant library for mutants unable to grow under specific
PT conditions and for identifying loci involved in pathogenicity,
PT comprises using signature tagged transposon mutagenesis -
XX
XX
PS Example 8; Page 147-148; 159pp; English.
XX
CC The present sequence is given in a specification relating to a method for
CC screening a library of mutants. The method comprises constructing a
CC library with insertions in genes and/or regulatory regions of the
CC organisms of interest, where the insertion contains a tag and/or a
CC transposon associated with a tag. The mutants are identified by
CC hybridisation of the tags to known sequences. The method is useful for
CC treating an individual suffering from a mycobacterial infection,
CC suspected of being infected with a Mycobacterium, or having been
CC exposed to an infectious Mycobacterium. It is also useful for
CC identifying and isolating mutants of actinomycetales and for identifying
CC compounds that have antibiotic activity. The method is used to identify
CC mutants of microorganisms, preferably an actinomycetales, such as
CC M. tuberculosis, M. bovis, M. leprae, M. avium, M. intracellulare and
CC M. paratuberculosis, that is unable to grow under specific conditions.
CC It is especially useful for identifying loci involved in pathogenicity.
CC It is useful in constructing vaccines. The method can be used to screen
CC multiple libraries concurrently. It can screen libraries of different
CC organisms or different strains of the same organism.
XX
SQ Sequence 6333 BP; 1010 A; 2010 C; 2202 G; 1111 T; 0 other;
XX
Query Match 50.9%; Score 51.4; DB 22; Length 6333;
Best Local Similarity 69.3%; Pred. No. 3; 1e-05;
Matches 70; Conservative 0; Mismatches 31; Indels 0; Gaps 0;
XX
OY 1 ATGCCCCGCCACTCGCGCGGAGATATTTCGACCGCTGTACCGCGAAGCGGCG 60
DB 4732 ATGCCCCGCCCAAGGAGGAGGAGATTTCGCGACCGCGCAATCCGCGAAGCGGCC 4791
OY 61 TGGCTGGCGGACGAGGATCGCGCACGTATGACTCGCG 101
DB 4792 ATGCTGCGCGACATGGCGCTGAGCATGTCTACGATTTCGCG 4832
XX
RESULT 7
AA199683/C
ID AA199683 standard; DNA: 4403765 BP.
XX
AC AA199683;
XX
XX
DT 15-JAN-2002 (first entry) ?
XX
DE Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 2.
XX
XX Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;
KW variation; epidemiology; patient treatment; epidemic monitoring; ds.
XX
OS Mycobacterium tuberculosis.
XX
PN US6294328-B1.
PD 25-SEP-2001.
XX
XX 24-JUN-1998; 98US-0103840.
PR 24-JUN-1998; 98US-0103840.
XX
XX (GENO-) INST GENOMIC RES.
XX
PI Fleischmann RD, White OR, Fraser CM, Venter JC;
XX
XX WPI: 2001-647261/74.
XX
PT Evaluating strain variation of Mycobacterium tuberculosis, comprises
PT determining the nucleotide sequence of the strain at positions in the
PT genome corresponding to positions where M. tuberculosis strains CDC
PT 1551 and H37Rv differ -

XX
XX Claim 4; SEQ ID NO 2; 3pp + Sequence Listing; English.
XX
PS
XX
CC The invention relates to evaluating strain variation within and between
CC different populations of the tuberculosis bacterial pathogen,
CC Mycobacterium tuberculosis or related Mycobacterium by determining the
CC nucleotide sequence of the first strain at positions in the complete
CC sequence of the genome that correspond to positions that differ in the
CC nucleotide sequences of M. tuberculosis strains CDC 1551 (AA199683) and
CC H37Rv (AA199682). The method is useful for evaluating strain variation of
CC M. tuberculosis and has valuable application in the fields of
CC tuberculosis genetics, epidemiology, patient treatment and epidemic
CC monitoring.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from USPRO
CC at seqdata.uspro.gov/sequence.html?DocID=6294328B1.
XX
SQ Sequence 4403765 BP; 757105 A; 1447799 C; 1441301 G; 757371 T; 189 other;
XX
Query Match 43.0%; Score 43.4; DB 22; Length 4403765;
Best Local Similarity 64.4%; Pred. No. 0.0062;
Matches 65; Conservative 0; Mismatches 36; Indels 0; Gaps 0;
XX
OY 1 ATGCCCCGCCACTCGCGCGGAGATATTTCGACCGCTGTACCGCGAAGCGGCG 60
DB 4287107 ATAGCCCGCGCGCGCGCTGAGATCTTCGCCACCGCGTACCCACAGCGTGAAG 4287048
OY 61 TGGCTGGCGGACGAGGATCGCGCACGTATGACTCGCG 101
DB 4287047 CTAATGCGCACATGGGTATTGAGCATGTCTATGACTCGCG 4287007
XX
RESULT 8
AA199682/C
ID AA199682 standard; DNA: 4411529 BP.
XX
AC AA199682;
XX
XX
DT 15-JAN-2002 (first entry)
XX
DE Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 1.
XX
XX Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;
KW variation; epidemiology; patient treatment; epidemic monitoring; ds.
XX
OS Mycobacterium tuberculosis.
XX
PN US6294328-B1.
PD 25-SEP-2001.
XX
XX 24-JUN-1998; 98US-0103840.
PR 24-JUN-1998; 98US-0103840.
XX
XX (GENO-) INST GENOMIC RES.
XX
PI Fleischmann RD, White OR, Fraser CM, Venter JC;
XX
XX WPI: 2001-647261/74.
XX
XX
DR WPI: 2001-647261/74.
XX
XX
PT Evaluating strain variation of Mycobacterium tuberculosis, comprises
PT determining the nucleotide sequence of the strain at positions in the
PT genome corresponding to positions where M. tuberculosis strains CDC
PT 1551 and H37Rv differ -
XX
XX Claim 3; SEQ ID NO 1; 3pp + Sequence Listing; English.
XX
XX
CC The invention relates to evaluating strain variation within and between
CC different populations of the tuberculosis bacterial pathogen,
CC Mycobacterium tuberculosis or related Mycobacterium by determining the
CC nucleotide sequence of the first strain at positions in the complete
CC sequence of the genome that correspond to positions that differ in the

OY 2 TCGCCGCGCACCTCGGCGGAGATATTTCGACCGCTGTACGCCGAGAACGGGCG 59
 DB 12650 TCGCAGCGCACCTCGGCGGAGGCTTTGCGACCGCCAGCTCCAGGAGAGTGAGCGC 12707

RESULT 11

ID AAT89956 standard; DNA: 28958 BP.

AC AAT89956;

DT 12-MAR-1998 (first entry)

DE Sorangium cellulosum soraphen gene cluster genomic DNA.

KW Soraphen; polyketide synthase; PKS; antipathogenic; antibiotic;

XX transgenic plant; phytopathogen; resistance; ss.

OS Sorangium cellulosum.

PN US5662898-A.

PD 02-SEP-1997.

PF 01-JUN-1995; 95US-0457342.

PR 20-AUG-1990; 90US-0570184.

PR 02-JUL-1992; 92US-0908284.

PR 31-AUG-1992; 92US-0937648.

PR 01-JUL-1993; 93US-0087636.

PR 08-JUN-1994; 94US-0258261.

XX (CIBA) CIBA GEIGY CORP.

PI Beck JJ, Gaffney TD, Hammer PE, Hill DS, Lam ST;

PI Ligon JM, Ryals JA, Schnupp T, Uknes SJ;

DR WPI; 1997-447901/41.

XX Protecting plants against pathogens with genetically transformed

PT biological control agent - which expresses all polypeptides)

PT Involved in pyrrolinitrin biosynthetic pathway

XX Ex 16; Column 104-128; 88pp; English.

PS This genomic DNA sequence encodes the Soraphen gene cluster. This

XX sequence encodes two open reading frames (ORF's), the positions of which

CC are not given in the specification. ORF1 is approximately 25.5 kb in

CC size and encodes 5 biosynthetic modules homologous to the erythromycin

CC genes of Saccharopolyspora erythraea. Each module contains a

CC beta-ketoadslysynthase (KS), an acyltransferase (AT), a ketoreductase

CC (KR) and an acyl carrier protein (ACP) domain. ORF2 is immediately

CC adjacent to ORF1 and is thought to contain 3 modules. Soraphen is a

CC type I polyketide synthase (PKS) which has antibiotic activity.

CC Transgenic plants containing such antipathogenic genes like those

CC encoded in the Soraphen cluster should have enhanced resistance to

XX attack by phytopathogens.

SO Sequence 28958 BP; 4084 A; 11863 C; 8356 G; 4655 T; 0 other;

Query Match 32.1%; Score 32.4; DB 18; Length 28958;

Best Local Similarity 72.4%; Pred. No. 5.3; Mismatches 16; Indels 0; Gaps 0;

OY 2 TCGCCGCGCACCTCGGCGGAGATATTTCGACCGCTGTACGCCGAGAACGGGCG 59

DB 12950 TCGCAGCGCACCTCGGCGGAGGCTTTGCGACCGCCAGCTCCAGGAGAGTGAGCGC 13007

RESULT 12

ID AAA75299 standard; DNA: 28958 BP.

XX

AC AAA75299;

XX 15-JAN-2001 (first entry)

XX DNA sequence of Sorangium cellulosum soraphen gene region.

XX Pyrrolinitrin; PrnA; PrnB; PrnC; PrnD; antibiotic; biocontrol;

XX fungal respiratory electron transport inhibitor; transgenic plant;

XX antipathogenic substance; biosynthetic gene; fungal resistance;

XX phytopathogen; soraphen; ss.

OS Sorangium cellulosum.

PN US6117670-A.

PD 12-SEP-2000.

PF 24-FEB-1998; 98US-0028934.

PR 30-MAY-1995; 95WO-IB00414.

PR 08-JUN-1994; 94US-0258261.

PR 09-OCT-1996; 96US-0729214.

XX (NOV) NOVARTIS FINANCE CORP.

PI Ligon JM, Lam ST, Hammer PE, Van Pee K, Kirner S, Young TR;

PI Hill DS;

DR WPI; 2000-593649/56.

XX Pyrrolinitrin biosynthesis genes and enzymes, useful for producing

PT pyrrolinitrin for increasing plant resistance to phytopathogens, e.g.

PT fungi or nematodes, and for developing microbial hosts useful as

XX biocontrol agents

PS Example 14; Column 121-146; 124pp; English.

XX The present sequence represents a Sorangium cellulosum soraphen gene

CC cluster. The specification describes pyrrolinitrin gene regions. This

CC region encodes enzymes required in the biosynthetic pathway of

CC pyrrolinitrin. The PrnA enzyme catalyses the conversion of tryptophan

CC to 7-chlorotryptophan. The PrnB enzyme catalyses the conversion of

CC 7-chlorotryptophan to monodechloroaminopyrrolinitrin. PrnC enzyme

CC catalyses the conversion of monodechloroaminopyrrolinitrin to

CC aminopyrrolinitrin. PrnD enzyme catalyses the conversion of

CC aminopyrrolinitrin to pyrrolinitrin. Pyrrolinitrin has antibiotic activity,

CC and inhibits a broad range of fungi. It is a fungal respiratory electron

CC transport inhibitor. The genes are useful for developing a host as a

CC biocontrol agent, for protecting plants against a pathogen and for

CC producing large, uniform amounts of pyrrolinitrin. The genes are also

CC useful for producing transgenic plants that express antipathogenic

CC substance (APS) biosynthetic genes. The genes and enzymes are useful

CC for increasing the resistance of host plants to disease caused by

CC phytopathogens, e.g. fungi, bacteria or nematodes. The microbial hosts

CC are useful for controlling or inhibiting the growth of a phytopathogen.

XX

SO Sequence 28958 BP; 4084 A; 11863 C; 8356 G; 4655 T; 0 other;

Query Match 32.1%; Score 32.4; DB 21; Length 28958;

Best Local Similarity 72.4%; Pred. No. 5.3; Mismatches 16; Indels 0; Gaps 0;

OY 2 TCGCCGCGCACCTCGGCGGAGATATTTCGACCGCTGTACGCCGAGAACGGGCG 59

DB 12950 TCGCAGCGCACCTCGGCGGAGGCTTTGCGACCGCCAGCTCCAGGAGAGTGAGCGC 13007

RESULT 13

ID AA05287 standard; DNA: 49377 BP.

XX AA05287;

XX

DT 21-MAY-1998 (first entry)
XX The soraphen biosynthesis gene cluster from Sorangium cellulosum.
DE
XX Polyketide synthase; PKS; biosynthesis; soraphen; SorA; SorB; SorB;
KW SORM; biosynthetic module; beta-ketoacylsynthase; acyltransferase;
KW ketoreductase; beta-ketone processing domain; cytosolic agent;
KW antimicrobial agent; phytopathogenic fungi; transgenic plant;
KW biological control; ss.
XX
OS Sorangium cellulosum.
XX
FH Key Location/Qualifiers
FT CDS 383..760
FT /tag= a
FT /product= SOR
FT /note= "gene product highly homologous to the
FT reductase domains of type I PKs such
FT as eryA from Saccharopolyspora erythraea"
FT CDS 927..19874
FT /tag= b
FT /product= SorA
FT /note= "gene product is highly homologous to
FT type I PKs that are known to be involved
FT in the synthesis of polyketide compounds"
FT misc_feature 942..7115
FT /tag= c
FT /note= "module 1 of SorA"
FT 7203..12884
FT /tag= d
FT /note= "module 2 of SorA"
FT 13455..19616
FT /tag= e
FT /note= "module 3 of SorA"
FT 19871..46318
FT /tag= f
FT 19871..46318
FT /tag= g
FT /product= SorB
FT /note= "gene product is highly homologous to
FT type I PKs genes"
FT misc_feature 19870..24556
FT /tag= h
FT /note= "module 1 of SorB"
FT 24638..30820
FT /tag= i
FT /note= "module 2 of SorB"
FT 30881..35446
FT /tag= j
FT /note= "module 3 of SorB"
FT 35528..40114
FT /tag= k
FT /note= "module 4 of SorB"
FT 40190..46318
FT /tag= l
FT /note= "module 5 of SorB"
FT 46851..47891
FT /tag= m
FT /product= SORM
FT /note= "gene product is homologous to the
FT methyltransferase from Streptomyces
FT hygroscopicus that is involved in
FT the synthesis of the polyketide rappamycin"
XX
PN US5716849-A.
XX
XX 10-FEB-1998.
PD
XX
PF 14-DEC-1996; 96US-0764233.
XX
XX 24-AUG-1993; 93WO-US07954.
PR 08-JUN-1994; 94US-0258261.
PR 09-OCT-1996; 96US-0729214.

XX
PA (NOVS) NOVARTIS FINANCE CORP.
XX Beck JJ, Hill DS, Liqon JM, Neff S, Ryals JA, Schupp T;
PI WPI; 1998-158369/14.
DR
XX
XX
PT DNA encoding Sorangium cellulosum polypeptide(s) - used for, e.g.
PT biosynthesis of soraphen useful as antimicrobial agent against
PT phytopathogenic fungi
XX
XX
PS Claim 2; Columns 47-90; 64pp; English.
XX
CC The present sequence contains a cluster of genes that encode polyketide
CC synthases (PKSs) that are involved in the synthesis of soraphen in
CC Sorangium cellulosum. The proteins encoded by the present sequence are
CC SorA, SorB, SorB and SORM. SorA and SorB contain biosynthetic modules
CC which contain a beta-ketoacylsynthase, an acyltransferase, a
CC ketoreductase and an acyl carrier protein domain, as well as beta-ketone
CC processing domains. S. cellulosum soraphen are useful as a cytostatic
CC and antimicrobial agent active against phytopathogenic fungi.
CC Soraphen-producing transgenic plants or biological control agents can
CC also be produced, which may reduce crop losses and nutritional
CC deprivation for local populations in many parts of the world.
XX
SQ Sequence 49377 BP; 7247 A; 19522 C; 14477 G; 8131 T; 0 other:
Query Match 32.1%; Score 32.4; DB 19; Length 49377;
Best Local Similarity 72.4%; Pred. No. 5.5;
Matches 42; Conservative 0; Mismatches 16; Indels 0; Gaps 0;
QY 2 TGGCCGCGCACCCTGGCGGAGATATTGGCAGCGCTGTCGCGGAGAGCGGCGC 59
DB 29253 TCGCAGCGCACCTCGCGCGGAGGTCCTTGGCCACGCGCAGTCCAGGAGTGGACGCG 29310
RESULT 14
AAAT78508
ID AAAT78508 standard; DNA; 44377 BP.
XX
AC AAAT78508;
XX
DT 26-FEB-1998 (first entry)
XX
XX Platenolide synthase gene cluster.
DE Platenolide synthase gene cluster.
KW Platenolide synthase gene cluster; platenolide production; srmg gene;
KW multi-functional protein; macrolide antibiotic; spiramycin; ss.
XX
OS Streptomyces ambofaciens.
XX
XX
FH Key Location/Qualifiers
FT CDS 350..14002
FT /tag= a
FT /transl_except= (pos:350..352, aa:Met)
FT /note= "ORF1 encodes protein shown in AAM23716"
FT 14046..20036
FT /tag= b
FT /note= "ORF2 encodes protein shown in AAM23717"
FT 20110..31284
FT /tag= c
FT /transl_except= (pos:20111..20113, aa:Met)
FT /note= "ORF3 encodes protein shown in AAM23718"
FT 31329..36071
FT /tag= d
FT /note= "ORF4 encodes protein shown in AAM23719"
FT 36155..41830
FT /tag= e
FT /note= "ORF5 encodes protein shown in AAM23720"
XX
PN EP791656-A2.
XX
XX 27-AUG-1997.

```

XX PF 19-FEB-1997: 97EP-0301066.
XX XX
XX PR 22-FEB-1996: 96US-0012050.
XX XX
XX PA (ELIL ) LILLY & CO ELI.
XX PI Bargett SG, Kuhstoss SA, Rao RN, Richardson MA;
XX PI Rosteck PR;
XX XX
XX DR WPI: 1997-418047/39.
XX DR P-PsDB: AAW23716-W23720.
XX PT DNA encoding Streptomyces ambofaciens platenolide synthase domain -
XX PT for production of spiramycin-related polyketide antibiotics
XX PS
XX PS Claim 9: Pages 8-33; 81pp: English.
XX CC
XX CC This sequence represents the platenolide synthase gene cluster of the
XX CC invention. This sequence is referred to as the smg gene, and was
XX CC isolated from Streptomyces ambofaciens. This sequence encodes the
XX CC multi-functional proteins which direct the synthesis of the polyketide
XX CC platenolide. Platenolide is the basic building block of the macroliide
XX CC antibiotic spiramycin. The DNA can be used to produce compounds,
XX CC including specifically the macroliide antibiotic spiramycin and spiramycin
XX CC analogues and derivatives. Modifications of the platenolide synthase DNA
XX CC sequence can be made so as to change the number and type of carboxylic
XX CC acids incorporated into the growing polyketide chain and to change the
XX CC kind of post-condensation processing that is conducted.
XX SQ
XX SQ Sequence 44377 BP; 4965 A; 15552 C; 17381 G; 6479 T; 0 other;
XX
XX Query Match 31.7%; Score 32; DB 18; Length 44377;
XX Best Local Similarity 68.2%; Pred. No. 7;
XX Matches 60; Conservative 0; Mismatches 25; Indels 3; Gaps 1;
XX
XX QY 1 ATGCGCCGCCACCTCGCGCGAGATATTTCGACCGCTGTCGCGGAGAAAGCGGCG 60
XX DB 29404 ATGCGCCGCCACCTCGCGCGAGAGTCTGCGCACCGC---GAGCCCGCGCAAGTGGGAC 29460
XX QY 61 TGGCTGCGGAGCAGGAGATCGCGCAGC 88
XX DB 29461 GCCCTGCGCGCATGGGCATCACCAGCG 29488
XX
XX RESULT 15
XX AAT80414
XX ID AAT80414 standard; DNA: 44377 BP.
XX AC AAT80414;
XX XX
XX DT 27-FEB-1998 (first entry)
XX DE Platenolide synthase gene cluster.
XX XX
XX KM Tyactone synthase gene cluster; tyIG gene; multifunctional protein;
XX KM platenolide synthase gene cluster; platenolide production; smg gene;
XX KM polyketide; tyactone synthesis; antibiotic; tylosin; ss.
XX XX
XX OS Streptomyces ambofaciens.
XX XX
XX FH Key Location/Qualifiers
XX FT CDS 350..14002
XX FT /*tag= b
XX FT /transl_except= (pos:350..352, aa:Met)
XX FT /note= "ORF1 encodes protein shown in AAW22606"
XX FT 14046..20036
XX FT /*tag= b
XX FT /note= "ORF2 encodes protein shown in AAW22607"
XX FT 20110..31284
XX FT /*tag= c
XX FT /transl_except= (pos:20111..20113, aa:Met)

```

```

FT FT /note= "ORF3 encodes protein shown in AAW22608"
FT FT 31329..36071
FT FT /*tag= d
FT FT /note= "ORF4 encodes protein shown in AAW22609"
FT FT CDS 36155..41830
FT FT /*tag= e
FT FT /note= "ORF5 encodes protein shown in AAW22610"
XX XX
XX PN EP791655-A2.
XX XX
XX PD 27-AUG-1997.
XX XX
XX PE 19-FEB-1997: 97EP-0301056.
XX XX
XX PR 22-FEB-1996: 96US-0012078.
XX XX
XX PA (ELIL ) LILLY & CO ELI.
XX XX
XX PI Dehoff BS, Kuhstoss SA, Rosteck PR, Sutton KL;
XX XX
XX DR WPI: 1997-418046/39.
XX DR P-PsDB: AAW22606-W22610.
XX XX
XX PT DNA encoding Streptomyces fradiae tyactone synthase domain - for
XX PT production of tylosin-related polyketide compounds
XX PS
XX PS Example 2: Pages 110-134; 220pp: English.
XX CC
XX CC This sequence represents the platenolide synthase gene cluster of the
XX CC invention. This sequence is referred to as the smg gene, and was
XX CC isolated from Streptomyces ambofaciens. This sequence encodes the
XX CC multi-functional proteins which direct the synthesis of the polyketide
XX CC platenolide. Platenolide is the basic building block of the macroliide
XX CC antibiotic spiramycin. This sequence was used along with the tyIG gene
XX CC (see AAT80413) to create a hybrid ORF1 sequence (see AAT80415). The tyIG
XX CC gene is the tyactone synthase gene cluster of the invention. The tyIG
XX CC sequence was isolated from Streptomyces fradiae, and encodes
XX CC multifunctional proteins which direct the synthesis of the polyketide
XX CC tyactone. Tyactone is the basic building block of the antibiotic
XX CC tylosin. The hybrid sequence can be used to transform S. ambofaciens
XX CC lacking the smg ORF1 sequence, or S. fradiae lacking the tyIG ORF1
XX CC sequence, so that they can produce polyketides. The DNA sequence can be
XX CC modified so as to alter the type of carboxylic acids incorporated, the
XX CC number of carboxylic acids incorporated and/or the post-condensation
XX CC reactions performed, thereby resulting in novel tylosin-related
XX CC polyketides.
XX SQ
XX SQ Sequence 44377 BP; 4965 A; 15552 C; 17381 G; 6479 T; 0 other;
XX
XX Query Match 31.7%; Score 32; DB 18; Length 44377;
XX Best Local Similarity 68.2%; Pred. No. 7;
XX Matches 60; Conservative 0; Mismatches 25; Indels 3; Gaps 1;
XX
XX QY 1 ATGCGCCGCCACCTCGCGCGAGATATTTCGACCGCTGTCGCGGAGAAAGCGGCG 60
XX DB 29404 ATGCGCCGCCACCTCGCGCGAGAGTCTGCGCACCGC---GAGCCCGCGCAAGTGGGAC 29460
XX QY 61 TGGCTGCGGAGCAGGAGATCGCGCAGC 88
XX DB 29461 GCCCTGCGCGCATGGGCATCACCAGCG 29488
XX

```

Search completed: November 6, 2002, 13:13:21
Job time : 1750 secs

OM nucleic - nucleic search, using sw model

Run on: November 6, 2002, 11:51:49 ; Search time 443.625 Seconds

(without alignments)
6625.829 million cell updates/sec

Title: US-09-724-876-2_COPY_36150_36250

Sequence: 1 atcgcgccacactcgcgc...gcgcacgtatgcactcgc 101

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

```
Minimum DB seq length: 0
Maximum DB seq length: 200000000000
```

Maximum DB seq length: 200000000000

Post-processing: Minimum Match 08

Listing first 45 summaries

Database :

GenEmbl.*
1: gb_da.*
2: gb_hg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vl.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mus.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*
29: em_vl.*
30: em_hg_hum.*
31: em_hg_in.*
32: em_hg_om.*
33: em_hg_mus.*
34: em_hg_pln.*
35: em_hg_rod.*
36: em_hg_mam.*
37: em_hg_vrt.*
38: em_sy.*
39: em_hgco_hum.*
40: em_hgco_mus.*
41: em_hgco_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	101	100.0	58733	1	AF217189	AF217189 Sorangium
2	101	100.0	71989	6	AF172664	AF172664 Sequence
3	99.4	98.4	68750	6	AF120843	AF120843 Sorangium
4	99.4	98.4	68750	6	AR193029	AR193029 Sequence
5	99.4	98.4	68750	6	AR199551	AR199551 Sequence
6	99.4	98.4	68750	6	AR199559	AR199559 Sequence
7	99.4	98.4	68750	6	AR199567	AR199567 Sequence
8	99.4	98.4	68750	6	AR201097	AR201097 Sequence
9	99.4	98.4	68750	6	AR208671	AR208671 Sequence
10	53	52.5	6462	6	AX153794	AX153794 Sequence
11	53	52.5	34071	6	AX153790	AX153790 Sequence
12	53	52.5	42717	6	AX153789	AX153789 Sequence
13	51.4	50.9	6333	6	AX069081	AX069081 Sequence
14	51.4	50.9	9699	1	MSCGM6ACSYN	M95808 Mycobacteri
15	51.4	50.9	14326	1	AE007123	AE007123 Mycobacte
16	51.4	50.9	24282	1	MYCY2AG1	283858 Mycobacteri
17	49.8	49.3	344050	1	MEPRPN1	AL583917 Mycobacte
18	48.8	48.3	42146	1	AE285636	AE285636 Burkholder
19	43.4	43.0	18160	1	AE007186	AE007186 Mycobacte
20	43.4	43.0	30352	1	MYCY409	297188 Mycobacteri
21	43.4	43.0	41171	1	U00010	U00010 Mycobacteri
22	43.4	43.0	332450	1	MEPRPN5	AL583921 Mycobacte
23	43	42.6	6219	1	AF340167	AF340167 Streptomy
24	42	41.6	16124	6	AX024384	AX024384 Sequence
25	42	41.6	16124	6	AX024277	AX024277 Sequence
26	39.6	39.2	41321	1	MSGJ409	AD000017 Mycobacte
27	37	36.6	11042	1	MYT045	AL022020 Mycobacte
28	37	36.6	19806	1	AE007024	AE007024 Mycobacte
29	35.4	35.0	14173	1	AE006999	AE006999 Mycobacte
30	35.4	35.0	37840	1	MYT005	AL010186 Mycobacte
31	34.6	34.3	10334	1	AE005083	AE005083 Halobacte
32	33.2	33.9	46166	1	SC8F11	AL135864 Streptomy
33	33.2	32.9	49736	1	AF319998	AF319998 Stigmatal
34	33.2	32.9	110000	2	LMFCH81B_09	Continuation (10 o
35	33	32.7	10477	1	AE012004	AE012004 Xanthomon
36	33	32.7	11930	1	AE004476	AE004476 Pseudomon
37	33	32.7	179837	2	AC127666	AC127666 Pseudomon
38	32.8	32.5	14980	1	SCA49	AL138668 Streptomy
39	32.6	32.3	107379	1	SHGCP1R	AL138668 Streptomy
40	32.6	32.3	192960	2	AC098808	X86780 S. hygroscop
41	32.6	32.3	306550	1	SM5591783	AC098808 Pabio cyn
42	32.4	32.1	3278	14	AF143953	AF143953 Sporozoa
43	32.4	32.1	14928	1	AE004676	AE004676 Pseudomon
44	32.4	32.1	28958	6	AR044578	AR044578 Sequence
45	32.4	32.1	28958	6	IA7768	IA7768 Sequence

ALIGNMENTS

RESULT 1	AF217189				
LOCUS	AF217189	58733 bp	DNA	linear	BCT 09-JUN-2000
DEFINITION	Sorangium cellulosum putative transposase gene, partial cds; putative transposase gene, complete cds; epothilone biosynthesis gene cluster, complete sequence; putative membrane protein gene, complete cds.				
ACCESSION	AF217189				
VERSION	AF217189.1	GI:7453554			
KEYWORDS					
SOURCE					
ORGANISM	<p> <i>Sorangium cellulosum</i>. <i>Sorangium cellulosum</i>. Bacteria; Proteobacteria; delta subdivision; Myxobacteria; Mycococcales; Sorangiineae; Polyangiacae; Polyangium. 1 (bases 1 to 58733) </p>				
REFERENCE					

AUTHORS Tang, L., Shah, S., Chung, L., Carney, J., Katz, L., Khosla, C. and
Jullien, B.
TITLE Cloning and heterologous expression of the epothilone gene cluster
JOURNAL Science 287 (5453), 640-642 (2000)
MEDLINE 20115953
PUBMED 10649995
REFERENCE 2 (bases 1 to 58733)
AUTHORS Jullien, B., Shah, S., Ziermann, R., Goldman, R., Katz, L. and Khosla, C.
TITLE Isolation and characterization of the epothilone biosynthetic gene
cluster from Sorangium cellulosum
JOURNAL Gene 249 (1-2), 153-160 (2000)
MEDLINE 20293058
PUBMED 10831849
REFERENCE 3 (bases 1 to 58733)
AUTHORS Jullien, B.
TITLE Direct Submission
JOURNAL Submitted (16-DEC-1999) Kosan Biosciences, Inc., 3832 Bay Center
Place, Hayward, CA 94545, USA
FEATURES
source Location/Qualifiers
1. 58733
/organism="Polysangium cellulosum"
/strain="SMP4"
/db_xref="taxon:56"
<1. 992
/note="orfA"
/codon_start=3
/transl_table=11
/product="putative transposase"
/protein_id="AAF62878.1"
/db_xref="GI:7453555"
/translation="VAGHEAFADFGCVPRVLLYDNLKNAVVERHGDAIRHPTLAL
SADYREPRVAVARGNEKGRVERAIRYREGFEKRAVADGLDLNRKQTEHTSNAL
DSRWEDRARTRQAFDDERSVLLRHPTDPFDEHREVEVEKTPARDLNLSYVH
DRTGRLVVLADLSQVRIADGNOIVATHYRSMRGOQIEOPHELQGVDEKRRARHR
GDLRLAARASQAFRLIYAEERGDVNSAIARLQDLDAVGAALEAEALVEYLERNTI
HTGAVQVDRRRERHRLPPPSIPTREHNAIVYTPHSITTYDLAKKDP"
999. 1501
/note="orfB"
/codon_start=1
/transl_table=11
/product="putative transposase"
/protein_id="AAF62879.1"
/db_xref="GI:7453556"
/translation="MTDLPTETKDKLNSGLGFLACMEQLADKRMFLREVALIEBRE
RHKRSLERLKNSRYAAEFKPMTPDFDSNPKRIDREAVDOLYDSRYADLFEVYTRDYD
AOKPLLSLNKAFADMGQVFPAAACVYTLVDRLVHRAVEIEAEISYRLKEAKELNAT
RTKORTKKH"
1998. 6263
/gene="epoA"
/gene="epoA"
1998. 6263
/gene="epoA"
/note="type I polyketide synthase; epothilone PKS loading
module"
/codon_start=1
/transl_table=11
/product="EpoA"
/protein_id="AAF62880.1"
/db_xref="GI:7453557"
/translation="MADRIERAEDPIAIVGAGCRLPFGVIDLSGFVTLGSRDVT
GOVPERMDAAAMPDPLDAPKCTPYTRASFISDVACFPASFGFISPPRALDGRDPAHR
LLEVCWEALENAIAPSAVLGTEGVFTIGIPSEYEAALPRTASAEIDAHGGLDSTM
PSVGAGRISYVLGRLPCVAVDTAVSSSLVAVHACQVYVLRSLSGEATLAGEVSLMISP
STYVWLSTKRALATDGRCAKFAVSEADGFGRGSCAVVYLRSGARADDRILAVIRG
SAIHDGASSGLTVNGSSQEIYLRKALADACCAASSVGVYEHGSGTLLDPIEIOA
LNAVIGLGRDVATPLLIGSVKTNLIGPEVAGSITGLIKVYLSTLHGOIIPAHHAALN
PRISMGDLRTYTRARTPMDNTPRRAGVSSFGSGTNAHYVLEAPAACTPPAPE
RRABDLVLSARTAAADAHARLRDHLFTPSQCGDVAFSLATRSAEHRGLVAAT
SSEGLRAALDAAGQOTPGVVYGIADSSRGKLAFLFTGQGAQTILGMRGLDVWPAF
REAFDLCYRLFNOELDRPLREYMAAEPASVDALADQTAFTQALFTPEYALAAVMS
WCEPELVAGHSIGELVACVAGVSELDVAVFLVARGRLMALPAGGMAVIAAPEA
DYAAVVAAPHAASVSTAAVNGPDQVYIAGAGPVAHIAAAMARGARTRKALHYSHAHS
PLMAWMLAFCGRVASEVSTRRSYLVSNLSKAGTIDEVSGCYVRRARVRYRRADS
VVALHAAGTVEVGPSTLGLVPAVLCDARPALLAASSRAGDEPATVLEALGLW

ANGLYSNAGLEPSSGGRVPLFTYMQGERTYITDKKADDAAGRRAPAGHDEYK
GAVKGGDRRSARLDHPPEPSGRREKVEEAGDPFLEIDEPEGDLRLVLRTERAPG
LGEVEIADRAAGLSFNDVOLALGAWPDLPGPNPLDILGEGCAGRIYAVGEGVGLV
VGOPIYLSAGAFATHTTSAALVYLRPOALSATPAAMPVYLYTAAVILADGIALOP
GERVLIHAATGGVGLAAVOMAHVGAENHATGATPEKARVLESLIGRVYSSRPFV
ADVRAATGEGEVYVNLNSGLDIDKSFNLRSHGRFVELCKRDCYAAONQGLRFLR
NLSFSLVDRGMLEPRARVRLFEELDLIAGVFTPIPIITLFIYARVADNFRMAQ
AOHLKLVLTLDPEVOIIRPIHAAGSGTGRDLDRLASAPARAALAEALRTQ
VSOVLRTEPIKGAELALFRLGMDSLIMAVEENRIRTEASIKLSTFTLSTPNIALLT
ONLDAIATSLSERVAENLRAGVQSDVFSGADQWEIIAL"
6260. 10492
/gene="epoB"
/gene="epoB"
6260. 10492
/gene="epoB"
/note="non ribosomal peptide synthetase; epothilone PKS
module 1"
/codon_start=1
/transl_table=11
/product="EpoB"
/protein_id="AAF62881.1"
/db_xref="GI:7453558"
/translation="MTINOLNELHEOGYKLAADERLOIOAPKALNPMLARISEH
KSTILMLRORLPASEIYDAPARERHVPPLDIOGSYWLGRGATVPSGIHAREYD
CTDLYARLSRAFRKYVARHMLRAHTLPDMQVYIEPYVDADIEITDRCGRRE
RLVSLRDAMSHRIYDTERPILYVAVRLDEQOTRLVSLDILINDLSLSTIEKDWL
SFYEDPETSLPYLELSYRDYVLALESRRKSEHQSMDYMKRWKRVLEPPMPKAD
PSTLEIRIRHTEOMLPDSMSRLKQVGERGLPTGYILAAFSVIGRMSASPEFTL
NITLNRPLVHPVRNDITGDFTSWLIDIDTDRDSFPORAKRIQEOLEMDHDDVS
GIEVOREARVYGIORGALFPYVLTSLNCOGVGYTSIORGTPLYSTIOTRPOILLDH
QIYEHDDGLVLAAMDYDGFPPDLDDMLAEVATLRLKLTREPNSEQRCSIPRAQLE
ARASNETLSLSEHTLGLFARVLEQLPRAVYSAKRTIYELLSRRLGKRLR
EOGARNTLVAAWMEKMGEOVVAIVLESGAAVYIDADIAPAEIHLHLHGEVGLR
LQPMIDGLKSPMPGILFVLDADVEGGLDLPMPQIOTSDLVLYYTGSGGLPK
GVMIHGRGVNLTIDINEREFVLDVLAATLSLSEFDSYDVDFGLIAGGIYVPA
SKTRDPAHMAALIEREKYTVMSVPALRISYVLEHGRPDLSARSLTSLSGMDIPV
GLPGELOAIRREVSYISLUGATREASINSTGYRYVRNVDSPKASIPGRRLNOTPYLD
EALREPVWPGQLYTIGVGLALGTWRDEKRNLSLVHPETGERLYITGDGRPLD
GNIEFMGRDNOIKLGRYVELGLEIETLKSHPNVADVIYVPGDAANKLLAVVP
EGTRRRAEODASLKTERYDARAHAAKADGSDGRVQFKLARHGLRDLGKPVVDL
TGLVREAGLDVYARRRSYRPLEAPIFPEVGRPLSCISVEPGRALPKERYPSAG
STYPOVTAAYAKSGRIEGVDEGFYYHPEFHLKLVSHGIERGAHVONDPDEAA
RGLLVGRIDAIESIXGSLSRFCLLAGYNAQLMEBARPSENTICVCGOPDEQVYR
PVLDRHSDVYVHGMGLKVPDRQVQCTLGDDSPKRAITRGAPGQDQHPADLEIR
PLRKLPEYMWPTVEVLDALPLTNGVYRKAALRERKDTSPRSHSGPRAADLLEI
LYAAVRYVLEGVLEVOLOOSFVLDGATSIHYVMRSLQKRLDREIATITELQYPMIGS
LASGVRBKSDDORPNMODRVEARRKGRRS"
10639. 16137
/gene="epoC"
/gene="epoC"
10639. 16137
/gene="epoC"
/note="type I polyketide synthase; epothilone PKS module
2"
/codon_start=1
/transl_table=11
/product="EpoC"
/protein_id="AAF62882.1"
/db_xref="GI:7453559"
/translation="MEDESSAIVAVIGSGRPPGARDLDEEFNRNLDOCTEAVQRESEQ
ELAAGCPALVLDPNYVAGVSLGDVDRFDPAFGISPREMELMDPOHRIPEMECAE
ALEMNGYPLTAEGSICVYACGNMSSYLTSLNHEPAPRMGMQOTLIGNDKLYAT
HYSTRNLNRPSTISYQKOSTSLAVNHLKMSLSDRECDMAAGCITVIRIHRAGTYV
AAGGIFSPGHCRAFDANANGITWNGGVVYLKPLDRLADLGDGVRAVILGSAATRV
GARKIGTFAPSGVGOQAIMEALIAVGEVARSIOYIEHNGTTLGDAIETRAARV
DDDASTRSCAIGSVKGTGHELSAAGLAGLKYVALLEHQLPSPLESPNSIDF
ASSPYVYTKDMWTGPPRAGVSPFGIGGTNAHVYLEAPAKLPAAPARSAL
FYVSAKSAALDAAARLRDHOAHQGISLGDVASTLTTSPMHRILMAAPSEAL
REGDLAARGOTPGAVRGCSPGNVPVYVFPQSGGVAMGQDLAAEERVHAAL
SACDRAIDAEAGWSLLAELADESSQLEIRDDVQVYFALAVAEALMRWGVGPDV
VIGHSGMEVAANVAGALSLEDAVAILICRSRLLRISGCEMAVTELSLEAEARLV
GYEDSVAVVSNPSRSTVLSGPAIGVLESLNKGVCYCRVAKVDVSHSDPOPLR
BDLLAALGGLRPAARAVMRSYTCAMVAGPELIGAMNNMLRQVRAEVVQAOIGP
CHGLFVEMSPHILITTSYEBNRRAAQNRAGVAGSIRKODERPMALTEGLAMQOIGP
VFMGRLPAGGRVPLPYTPWORERTYIEAPKASAGGRVAGGHPLLGEMQTLST

QSTRMWTETLDELKRLPWLGDHRYGVAVPEPGAYLEMAISSGAEALGD;PLOTIDV
LAEALAFAGDAVAVYVTTTEOPSGRLOFOLASRPGGHSFPHHAGALRVERTE
VPAGITLSAVRARILOASMPAAATYAEITMELOYQAFNOGIAELMRGEGALRVERTE
DAAGSAAYRHLRHPALIDNACPOYVSLPAGGEGATPWPVEVGSRLRLOPSEGLMCHA
RVNHGRQTPDROGADENVDSGAVVAEVSGLVAQRI;PGVSVRREEDMDLMEWP
AAVGTAVNAGRMWLLGGGGGLGALRSMEGAAVHAHVESRTTSAAGVALLAKAP
DGOPTAVVHLGSLDGGELDPGLGAGALDAPRSADVSPALDPALVRCGDSVLMTP
OALAGMPERDAPRLMLTLTRGAQVAGAGVSTQAPLGLGRTVIMEHEDLCARVLD
PTRPDSIGLALLIADDAEVALRNGEFCVARIYRORPETERPREGIESCPDWT
IRADSTVYTGGLGLGLSVAGMLAEKAGCHLYVYGSNGASVQDRAVALLEKAR
VTAKADVADRAQLERILREVTTSGMPLRGVNAAGLIDDLMOQTPARRKVMAP
VOGALHJHALTREALSEFVLYASVGLSGGQNTAAATFTLDALHHRAGGLPA
LSVDMGLTAEVMAAQAQEDRGARLVSRKMRSLTPDEGLSALRLESSEGVYGLPV
PRLWELYPAAASSRMLSLVTAHARSAGPAGDGLLRRLAAAPESAFSGLEPLLR
AOISOVLRLPGKLEVDAPLTLGMSNLMGLIELRNRIEAMGITVPATLMTYPTVA
LSGHLARACCAAAVESPHTTADSANVELIEMSODLITOLIAKRAKALT"
16134. .37907
/gene="epod"
/gene="epod"
16134. .37907
/gene="epod"
/note="type 1 polyketide synthase; epothilone PKs modules
3, 4, 5, and 6"
/codon_start=1
/transl_table=1
/product="epod"
/protein_id="AAF62883.1"
/db_xref="GI:7453560"
/translation="MTTRGPTAQNPLKQAIIITQRLERLAGLAQALETEPIATV
GIGRFGGADAPAEFWELLDABEDAVQPLDRNALVGAVEAVPHHAGLTPEDIC
FDAEFGISPREARSLDPORILLLEVAMEGLDAGIPRSIDGSRGTGFEVAFADYA
RTVARLPRERDAYSATGNMLSLAAGRLSYLGGPCLTYDTCSSVLVLIHLACS
LRAGESDLAAGVSTLSPDMEMAAARTOLSPDRCRTFDACNGSVREGGGIVY
LKRLSDAORODRIWALIRGSAIINHDSRTGLPAPNLAQCTYVIREAIRSHVAGAY
DYVTRGTGTSIDGPIEYERLKVATVGPARGSGTCVGAFTNIGHLPAAGVGLIK
AALSLTHERIPRNINFTLRNPRIRLESALALATEPVWPTDPRRAGVSSFGMSGT
NAHVLEAPAEVLEMPAPAPERSAELVLSGSGEALDPAQAARLHMLHMLGLGVY
AFSLATRSATVHRLAAVTSREGILVLSVAGGOLPAGARCIASISRGKLLFT
GOGAOTPGMGKGLCAAMPAPAREARCCVTELDRLDRLRPAHMAFASARSLLDQ
AFQOPALFAVRYATATLAWRMSGVPELLVGHISIEYLVAACVAFPSLEDRVLYAAR
RLMGSLSGGAMWSLGAPEAEVAAVAAPHAWSVIAAVNGEDVYVIAVEAVOALAA
GFAARGVKTRKLHVAHFHSPLMPEMLEEFCRVAASYTRRPSVLSVNSLKGKVTDE

Query Match 100.0%; Score 101; DB 1; Length 58733;
Best Local Similarity 100.0%; Pred. No. 1.5e-14;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATCGCCCGCCACCTCGGCGGAGATATTGGACCGCTGTGACCGCGAGAGAGCGGCG 60
|||||
Db 36150 ATCGCCCGCCACCTCGGCGGAGATATTGGACCGCTGTGACCGCGAGAGAGCGGCG 36209
|||||

Qy 61 TGGCTGCGCGAGCAGGAGATCGCGACGTGATGACTCGCG 101
|||||
Db 36210 TGGCTGCGCGAGCAGGAGATCGCGACGTGATGACTCGCG 36250
|||||

RESULT 2
AR172664
LOCUS AR172664 71989 bp DNA linear PAT 17-DEC-2001
DEFINITION Sequence 2 from patent US 6303342.
ACCESSION AR172664
VERSION AR172664.1 GI:17912155
KEYWORDS
SOURCE .
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 71989)
AUTHORS Julien,B., Katz,L., Rhosla,C. and Tang,L.
TITLE Recombinant methods and materials for producing epothilones C and D
JOURNAL Patent: US 6303342-A 2 16-OCT-2001;
FEATURES
Location/Qualifiers
1. .71989
/organism="unknown"
BASE COUNT 10108 a 23531 c 26617 g 11731 t 2 others

ORIGIN
Query Match 100.0%; Score 101; DB 6; Length 71989;
Best Local Similarity 100.0%; Pred. No. 1.4e-14;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATCGCCCGCCACCTCGGCGGAGATATTGGACCGCTGTGACCGCGAGAGAGCGGCG 60
|||||
Db 36150 ATCGCCCGCCACCTCGGCGGAGATATTGGACCGCTGTGACCGCGAGAGAGCGGCG 36209
|||||

Qy 61 TGGCTGCGCGAGCAGGAGATCGCGACGTGATGACTCGCG 101
|||||
Db 36210 TGGCTGCGCGAGCAGGAGATCGCGACGTGATGACTCGCG 36250
|||||

RESULT 3
AF210843
LOCUS AF210843 68750 bp DNA linear BCT 21-JAN-2000
DEFINITION Sorangium cellulosum strain So ce90 epothilone biosynthesis gene
cluster, complete sequence.
ACCESSION AF210843
VERSION AF210843.1 GI:6724237
KEYWORDS
SOURCE .
ORGANISM Polyangium cellulosum.
Polyangium cellulosum.
Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
Myxococcales; Sorangineae; Polyangiaceae; Polyangium.
Mojnar,I., Schupp,T., Ono,M., Zitzke,R., Mlhamow,M.,
Nowak-Thompson,B., Engel,N., Touzet,C., Strattan,A., Cyr,D.D.,
Gorlach,J., Mayo,J.M., Hu,A., Goff,S., Schmid,J. and Ligon,J.M.
The biosynthetic gene cluster for the microtubule-stabilizing
agents epothilones A and B from Sorangium cellulosum So ce90
10662695
20130945
PUBMED 10662695
REFERENCE 2 (bases 1 to 68750)
AUTHORS Mojnar,I.
TITLE Direct Submission
JOURNAL Submitted (03-DEC-1999) Natural Product Genetics, Novartis
Agricultures Research Institute, Inc., 3054 Cornwalis Rd, P.O.Box
12257, Research Triangle Park, NC 27709, USA
FEATURES
Location/Qualifiers
1. 68750
/organism="Polyangium cellulosum"
/strain="So ce90"
/db_xref="taxon:56"
<1. .1826
/note="Orp1"
/codon_start=3
/transl_table=1
/product="unknown"
/protein_id="AAF26904.1"
/db_xref="GI:6724238"
/translation="ASIDALFARATSAVRLDDHGRAPRHHVLAENGIDRLALREH
LRIDGGSEFRCMKLGDITVFLAHNDPLASISHHARSLRHPWMTYADVPALY
RWLARGAPGLREYEBERARTAOERRLWLAAPCPAPDLPREEDDNGPLLP
MSPEVAERLRASVATPELACALALAWLTAGTGPMSGPAYEMLEPLTYAMGNADN
LISFEREALIAEVRRAOPAPPAAGGLAVAGVSSGRLSGVTDGALYSGDND
LIMFOPRISIPVVLVAGNDPEFLAPLSQMLFAHNAGTISKVLTEGSLPYMANR
QARMSLYHARGFAMVNOAMVPDEGAPRVYORSTIMEPHEPTPCILPAGSARS
LACDEHLYMCELSAGRIELMRHPRHHRGAFSRAYIGEHPITATWYPSLTNATHYL
WADPDRAITLGVDRGTVEPRLVLAETLHHPAHVSESDNDIFALTGDPDSRWHEHIL
SGASTVADYORQLMDRDPDVLNRRGLFFITNDRILYARS"
complete(1900. .3171)
/note="Orp2: similar to hypothetical proteins from
Mycobacterium tuberculosis and streptomyces coelicolor and
to DD peptidases"
/codon_start=1
/transl_table=1
/product="unknown"
/protein_id="AAF26905.1"

/db_xref="GI:6724239"
/translation="MGALISVAAPGALGAEDEEGOPGODAGAGALAPAREVMAAEVA
ACQMPGAVMLVARGDDVDVADAVTELGSGAPMRDITFRIASMTKAVATAMVME
EGKLDDSPEDRLPELARNKRVLRIDGPIDETIVRPIITVMDLMTFMGEGISDA
SSPIORAIIDELGLVNAOPVMPHPGDEMRIRGCTPIHQPGACOMMYNTGSGVCL
VGAADQGFDAVYRERILAPLGRDIDFHVPPADKLRFAGCCGFTDEQTEGKTRMD
GASAVASPPAPFSGAAGLVSTVDYLLFRMLMNGVHEGRLLISAASREMTACHL
TPAQKASSFPFPGFTHGMCMAVVTAPDAVEVPGRGYMGDGGFTSMINDPGREL
IGIVMTOSAGFLFSGALERFMRSVYATESA"
3415..5556
/note="ORF3; similar to Na/H and drug/H antiporters"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAF26906.1"
/db_xref="GI:6724240"
/translation="MHGTEROVLISVLTALITVYARASGEIARLRPEVIGELFG
GVYLGPSVVGALPGFHRALFOEPAGVVLSTIGALLLLMAGIEVDGILREEA
RPGALSALGAIAPLPAAGAFSALVLDRLPSGLFGIVSVYAVIAKVLTERESM
RSYAGVTLAAGVSEVAAVLAAMTSSSGASPALAVARSALASGFLFMVLRGR
LTHLARWVADATRVSGQVSLVLTFLAALTORLGHPLGAFALGVLVAVSAPRT
NRLDGVOTIVAGLFAPVFEVLAGMRVDSOLRTPAANGTVALLATPTAKAVVAA
LGARLGLGSEALIVAGLNMKGTDILVATYGVGLISNATIMYAVVAVLYTA
SPALIMLEKRAPPTIOESARLEKEBAARATIPGERILVPIVAHALGFATDIVES
IVASKRKLETVDTIELSVYEQAPGPSRAGEASNLARLGRVGIWRORRELGS
IOAILRASRDHDLVIGARSPARAGMSFGLDDAIVORAESNVLVGDPPABERAS
ARILVPIIGLESLFAADLAHVLAAMDALVLISSAOTDPCAIVVMRREPVRVAV
ARSVDVAEVRGRGLGVSRVYHVGAFHSPDETRRLARAPLDLVLCGYDHGPIGLRL
YLGSTVESYVRSRVPALLVHAGTREQVR"
complement(5612..5992)
/note="ORF4"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAF26907.1"
/db_xref="GI:6724241"
/translation="MDKPIGTRCAIAEYIPEGSGPEPQMTSHEACILNANSRDA
QVATIVYFSDRDPAGPYRVTVPARKTRHVRFDLTPEPIPRDTYASVIESDAPIV
OHLRLDSRGAENMLSTIAYDRE"
6226..6675
/note="ORF5"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAF26908.1"
/db_xref="GI:6724242"
/translation="MKHVDVGRFRGRIRGHTIGLASMALAGCGGSEKTVGTRAP
GADARTADVDPDAATTRLAVDVYHLSPPERLEAGSERVYVMQRSPSPWRVGYLD
YNADSRGKLAETVYANFELLITAEKSSPOSPSSAIVIGTYSV"
7610..11875
/gene="epoa"
/gene="epoa"
7610..11875
/gene="epoa"
/note="EPOS A"
/codon_start=1
/transl_table=11
/product="polyketide synthase"
/protein_id="AAF26919.1"
/db_xref="GI:6724253"
/translation="MADRIERAEDPIAIVGASCHLPBGVIDLSGFTLLGSRDY
GVPAERMDAAMFDDPDAPGKTPTRASFUSDVACFPDASFEGISPRALRMDPXR
LLEVCWEALENNALAPSAVSTETVEFGIPSEVEALPQATASAEIDAGGLTGM
PSVAGRIAYALRGPCVAVDTAYSSSLVAHLACQSLRSECSALAGVGLIMSP
STLVMLSKTRLARDRGCAKFASEADGFGRGCAVYVLRSGARGLDRIIAYVR
SAIINHGAASSGLJVPNGSSOEIVIKRALADAGCAASGVYEAHGTGTDPIETQIA
LNAVYGLGDVATPLLIGSKYTNLGRPEASGITGLKAVLSIQHGQITAHILHAQALN
PRIISWGLDLVTYRATPWPDMNTPRRAGVSSFGMSGTNAHVLEAPATCTPPAPE
RPAELVLSARTASALDAQARLRDHLERYPSOCGLDVAFSLATRSAMEHRLAVAT
SREGLLAALDAAGQOTSPGAVSIADSSRKLAFLEFGQAGTQLMGGLVDVMAF
REAFDLVRLFNOELDRPLREYMMAPASVDAILDQTAFTQPALTFEYALAAWMS
KGVEPELVAGHSIGELVACVAGVSTEDAVFLVARGRLMALPAGCAVMSIEAPEA
DVAAAVAPHAASYSTAAVAPQOYVYIAGAGQFVAITAAAMARGARTKALHSHAFFS
PLMAPLEAFGRVAESVYRSPSYLVSNLSGKACTDEVSPPGYWVRHARLEVVRADG

VKALHAAGATEVENGPKSTLLGLVPACMPDARPALLASSRAGDEPATVLEBALGLM
AVGGLVSNAGRLPPSGGRVPLPTTYMQRERXYIDRKADDAAGDRAGACVDEEAG
GAVRGDDSEARGLDHPPESSGREKREKVAEDPFLRIEDEPVLVLRVVERAPG
LGEVEIYAVNAGLSENDVOLAGMPDDLRGKNDLPLGGSCACARIYAVGCVAGIV
VQOPIYALNAGCAFHYHTVSALVLPVRQALSAITAAAMPYALITVYALDITAILOP
GERVLIHAATGAGVGLAIVOMACHVGEVATAGTEKRAYLDSLVRYVSDSRQFV
AVVRAMTGGEEVDVYVNLISGRLIKSFNLSHRFELGRDCAVANOGLRFLR
NISFSYLDJRGKMLREPARVRLLELLELLIAGVTFPPPIATPLIARVADAFRMAQ
AOHLGKVLVLTIDPEVOIRIPTHAGAGPSTGDRDLRLASAPARAALAEAFIATO
VSOUVRPEIKVGAELFTRLGMDSLMAYELRNRIEASIKIKLSITTFIISTSPNMLLA
QNLIDALATALSIERVAENLNAQVNDPVSSGAQODWEITAL"
7640..8920
/gene="epoa"
/note="Region: beta-ketoacyl synthase/malonyl-CoA
decarboxylase"
9236..10201
/gene="epoa"
/note="Region: acetate acyltransferase"
10529..11428
/gene="epoa"
/note="Region: enoyl reductase"
11549..11764
/gene="epoa"
/note="Region: acyl carrier protein"
11872..16104
/gene="epop"
11872..16104
/gene="epop"
/note="EPOS P; contains condensation/meterecyclization,
cysteine adenylation, putative NAD(P)H oxidase, and
peptidyl carrier domains"
/codon_start=1
/transl_table=11
/product="nonribosomal peptide synthetase"
/protein_id="AAF26925.1"
/db_xref="GI:6724259"
/translation="MTINOLNLELHGCIKLADGERLOIQAPKALNPMLLARISEH
KSTILMLNORLPASIVAPAREHAPPLDIOESYWGRTGATVPSGIHAYREYD
CTDLVPLRSARAFKRVYARHDMLRHNTLPDMQVTEPKVADIEITIDGLDRSTREA
RLVLRSDAMSHRIYDTERPLVAVVAVRDEQOTRLVSTIDLINVDLSLSTIFKMD
SPYEDPESLPIVLELSYRDYVALLESRSKSEAHOSMDYKRIAEILPPEPLPKAD
PSTIKETIRFRHTEQMLPDSGKRLKRRGECRLTPTGYILAFSEYITGKWSAPFTL
NITLFRLEPHRPVANDITGDTSMVLDDITDTRDSFGQARARIEQMLEAHDCHVS
GTEVOREARVHIGQALFEPVYLSALNOQVGVTSLORETGPYVSTQTPQLDHD
QYEHGDGLVLAAMDIVDGVFPVLDLDMLEAAVVLRLRLETPMGEQVSCSLPPAOLE
ARASANATNALLSEHTLHGIFAARVLOLMLEAVVARSARTILEELSSRSRLGARLR
EGCARPNTIVAVMEKMGQVOYAVIAVLESGAAVYDADLPAERIHVLDHGEVKIV
LTPQMDGLKSLMPPTGIQRLVSEAVYEBGDGDPMPMPITQTPSDLAIVYITSGSTGLPK
GVYIDHRGAVNTILIDINERFETGPDQVYALSSLSFSDSYDVEGLIAGGTVVDA
SKLRDPAHMAELIEREKYVYVNSVPALMKMLVEHREGSPDSLARSRLSLSGMDIPIV
GLPGELOAIRPGVSVISLGATASISIGYVVRVADLSMASIPYGRPLRNOQTFVLD
BALPERPVWPGOILYIGVGLAGWRDEKRSSELVHPETGERLTYKTGDSGRYLPD
GNIEFMGRDNOIKRGYRVLEGETIEETLKSHPNRYDAIVPYVGDAAANKILLAAVVP
EGTRRRAAODASLITERIDARNAHAEBDGLSDGRVOPFKLARHGLRDLDBSKPYVD
TGQDPREAGLDVYARRRSYRTLEAPIPVERGRFLSCISVPEPGATLPKFRPSAG
STYPQOTYAVYKSGRIEGVDEGFYVYHPEGRFLLSHGIERGAHVNONDVEDPEA
FNLLEPGRIDAIESLYGSSREFCLLEGYAMOLMEAPQSLNIVCVQGFNEQVOR
PVLDLRSHDVYVYGMGLGVRDROVOVCTLGODSSPRATTPGAPRGEOHDFMLDRD
FLRKLTPYMWPTVFEVLEDAIPLTNSGVDRKALRERDTSPPRSRGHTAPADLLEI
LVAVYREVYGLSEVGLQOSFVDLGATSHIYMKSLKRLDRREIATITELTQYPLGGS
LASGLRDSRDLDOQRNMQDRVRRKGRRS"
16251..21749
/gene="epob"
/gene="epob"
16251..21749
/gene="epob"
/note="EPOS B"
/codon_start=1
/transl_table=11
/product="polyketide synthase"
/protein_id="AAF26920.1"
/db_xref="GI:6724254"

Query Match

98.4%; Score 99.4; DB 1; Length 68750;

Best Local Similarity 99.0%; Pred. No. 3.4e-14;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATCGCCCGCCACCTCGGCGCGAGATATTGGACCGCTGTGTACACCGAGAACGGGGCG 60
|||||
Db 41762 ATCGCCCGCCACCTCGGCGCGAGATATTGGACCGCTGTGTACACCGAGAACGGGGCG 41821
|||||

QY 61 TGGCTGCGGAGACAGGGGATCGCGACGTGATGACTCGCG 101
|||||
Db 41822 TGGCTGCGGAGACAGGGGATCGCGACGTGATGACTCGCG 41862
|||||

RESULT 4
ARI93029 68750 bp DNA linear PAT 20-APR-2002
LOCUS
DEFINITION Sequence 1 from patent US 6346404.
ACCESSION ARI93029
VERSION ARI93029.1 GI:20238994
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
SOURCE
BASE COUNT 9596 a 22456 c 25539 g 11159 t
ORIGIN

Query Match 98.4%; Score 99.4; DB 6; Length 68750;
Best Local Similarity 99.0%; Pred. No. 3.4e-14;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATCGCCCGCCACCTCGGCGCGAGATATTGGACCGCTGTGTACACCGAGAACGGGGCG 60
|||||
Db 41762 ATCGCCCGCCACCTCGGCGCGAGATATTGGACCGCTGTGTACACCGAGAACGGGGCG 41821
|||||

QY 61 TGGCTGCGGAGACAGGGGATCGCGACGTGATGACTCGCG 101
|||||
Db 41822 TGGCTGCGGAGACAGGGGATCGCGACGTGATGACTCGCG 41862
|||||

RESULT 5
ARI99551 68750 bp DNA linear PAT 20-APR-2002
LOCUS
DEFINITION Sequence 1 from patent US 6355457.
ACCESSION ARI99551
VERSION ARI99551.1 GI:20249625
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
SOURCE
BASE COUNT 9596 a 22456 c 25539 g 11159 t
ORIGIN

Query Match 98.4%; Score 99.4; DB 6; Length 68750;
Best Local Similarity 99.0%; Pred. No. 3.4e-14;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATCGCCCGCCACCTCGGCGCGAGATATTGGACCGCTGTGTACACCGAGAACGGGGCG 60
|||||
Db 41822 TGGCTGCGGAGACAGGGGATCGCGACGTGATGACTCGCG 41862
|||||

Db 41762 ATCGCCCGCCACCTCGGCGCGAGATATTGGACCGCTGTGTACACCGAGAACGGGGCG 41821
QY 61 TGGCTGCGGAGACAGGGGATCGCGACGTGATGACTCGCG 101
|||||
Db 41822 TGGCTGCGGAGACAGGGGATCGCGACGTGATGACTCGCG 41862
|||||

RESULT 6
ARI99559 68750 bp DNA linear PAT 20-APR-2002
LOCUS
DEFINITION Sequence 1 from patent US 6355458.
ACCESSION ARI99559
VERSION ARI99559.1 GI:20249633
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
SOURCE
BASE COUNT 9596 a 22456 c 25539 g 11159 t
ORIGIN

Query Match 98.4%; Score 99.4; DB 6; Length 68750;
Best Local Similarity 99.0%; Pred. No. 3.4e-14;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATCGCCCGCCACCTCGGCGCGAGATATTGGACCGCTGTGTACACCGAGAACGGGGCG 60
|||||
Db 41762 ATCGCCCGCCACCTCGGCGCGAGATATTGGACCGCTGTGTACACCGAGAACGGGGCG 41821
|||||

QY 61 TGGCTGCGGAGACAGGGGATCGCGACGTGATGACTCGCG 101
|||||
Db 41822 TGGCTGCGGAGACAGGGGATCGCGACGTGATGACTCGCG 41862
|||||

RESULT 7
ARI99567 68750 bp DNA linear PAT 20-APR-2002
LOCUS
DEFINITION Sequence 1 from patent US 6355459.
ACCESSION ARI99567
VERSION ARI99567.1 GI:20249641
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
SOURCE
BASE COUNT 9596 a 22456 c 25539 g 11159 t
ORIGIN

Query Match 98.4%; Score 99.4; DB 6; Length 68750;
Best Local Similarity 99.0%; Pred. No. 3.4e-14;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATCGCCCGCCACCTCGGCGCGAGATATTGGACCGCTGTGTACACCGAGAACGGGGCG 60
|||||
Db 41762 ATCGCCCGCCACCTCGGCGCGAGATATTGGACCGCTGTGTACACCGAGAACGGGGCG 41821
|||||

QY 61 TGGCTGCGGAGACAGGGGATCGCGACGTGATGACTCGCG 101
|||||
Db 41822 TGGCTGCGGAGACAGGGGATCGCGACGTGATGACTCGCG 41862
|||||

RESULT 8
LOCUS AR201097 68750 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 1 from patent US 638719.
ACCESSION AR201097
VERSION AR201097.1 GI:20251985
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and Gorlach,J.
TITLE Genes for the biosynthesis of epothilones
JOURNAL Patent: US 638719-A 1 19-MAR-2002;
FEATURES
source Location/Qualifiers
1..68750
/organism="unknown"
BASE COUNT 9596 a 22456 c 25539 g 11159 t
ORIGIN
Query Match 98.4%; Score 99.4; DB 6; Length 68750;
Best Local Similarity 99.0%; Pred. No.3.4e-14;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Query 1 ATGCCCCGACCTCGCGCGGAGATATTGGCAGCGCTGTACCGCGGAGAGCGGCG 60
Db 41762 ATGCCCCGACCTCGCGCGGAGATATTGGCAGCGCTGTACCGCGGAGAGCGGCG 41821
Qy 61 TGGCTGCGGAGCAGGAGATCGCGCAGCTGATGACTCGCG 101
Db 41822 TGGCTGCGGAGCAGGAGATCGCGCAGCTGATGACTCGCG 41862
RESULT 9
LOCUS AR208671 68750 bp DNA linear PAT 20-JUN-2002
DEFINITION Sequence 1 from patent US 6383787.
ACCESSION AR208671
VERSION AR208671.1 GI:21509886
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and Gorlach,J.
TITLE Genes for the biosynthesis of epothilones
JOURNAL Patent: US 6383787-A 1 07-MAY-2002;
FEATURES
source Location/Qualifiers
1..68750
/organism="unknown"
BASE COUNT 9596 a 22456 c 25539 g 11159 t
ORIGIN
Query Match 98.4%; Score 99.4; DB 6; Length 68750;
Best Local Similarity 99.0%; Pred. No.3.4e-14;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Query 1 ATGCCCCGACCTCGCGCGGAGATATTGGCAGCGCTGTACCGCGGAGAGCGGCG 60
Db 41762 ATGCCCCGACCTCGCGCGGAGATATTGGCAGCGCTGTACCGCGGAGAGCGGCG 41821
Qy 61 TGGCTGCGGAGCAGGAGATCGCGCAGCTGATGACTCGCG 101
Db 41822 TGGCTGCGGAGCAGGAGATCGCGCAGCTGATGACTCGCG 41862
RESULT 10
LOCUS AX153794 6462 bp DNA linear PAT 22-JUN-2001

DEFINITION Sequence 118 from Patent WO0140497.
ACCESSION AX153794
VERSION AX153794.1 GI:14535430
KEYWORDS
SOURCE unidentified bacterium.
ORGANISM Bacteria; environmental samples.
REFERENCE 1 (bases 1 to 6462)
AUTHORS Jeanlin,P., Perinotet,J.L., Guerinneau,M., Simonet,P., Courtols,S., Cappellano,C., Francou,F., Raynal,A., Ball,M., Sezonov,G., Tuphile,K. and Frostegard,A.
TITLE Method for obtaining nucleic acids from an environment sample, resulting nucleic acids and use in synthesis of novel compounds
JOURNAL Patent: WO 0140497-A 118 07-JUN-2001;
Aventis Pharma S.A. (FR)
FEATURES
source Location/Qualifiers
1..6462
/organism="unidentified bacterium"
/db_xref="taxon:2338"
BASE COUNT 1073 a 2067 c 2135 g 1187 t
ORIGIN
Query Match 52.5%; Score 53; DB 6; Length 6462;
Best Local Similarity 70.3%; Pred. No.0.0042;
Matches 71; Conservative 0; Mismatches 30; Indels 0; Gaps 0;
Query 1 ATGCCCCGACCTCGCGCGGAGATATTGGCAGCGCTGTACCGCGGAGAGCGGCG 60
Db 4750 ATGCGACAGCGTGGCGCGGAGATCTTGGCTACCTCCGCGGAGTCCGGAACAGAGCG 4809
Qy 61 TGGCTGCGGAGCAGGAGATCGCGCAGCTGATGACTCGCG 101
Db 4810 TATCTGCGCTGCTGGGATCGGCATGTTCGATTCGCG 4850
RESULT 11
LOCUS AX153790 34071 bp DNA linear PAT 22-JUN-2001
DEFINITION Sequence 114 from Patent WO0140497.
ACCESSION AX153790
VERSION AX153790.1 GI:14535426
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct.
REFERENCE 1 (bases 1 to 34071)
AUTHORS Jeanlin,P., Perinotet,J.L., Guerinneau,M., Simonet,P., Courtols,S., Cappellano,C., Francou,F., Raynal,A., Ball,M., Sezonov,G., Tuphile,K. and Frostegard,A.
TITLE Method for obtaining nucleic acids from an environment sample, resulting nucleic acids and use in synthesis of novel compounds
JOURNAL Patent: WO 0140497-A 114 07-JUN-2001;
Aventis Pharma S.A. (FR)
FEATURES
source Location/Qualifiers
1..34071
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="insert d'ADN du cosmid a2661 - brin codant"
BASE COUNT 5791 a 10858 c 11089 g 6333 t
ORIGIN
Query Match 52.5%; Score 53; DB 6; Length 34071;
Best Local Similarity 70.3%; Pred. No.0.0036;
Matches 71; Conservative 0; Mismatches 30; Indels 0; Gaps 0;
Query 1 ATGCCCCGACCTCGCGCGGAGATATTGGCAGCGCTGTACCGCGGAGAGCGGCG 60
Db 22973 ATGCGACAGCGTGGCGCGGAGATCTTGGCTACCTCCGCGGAGTCCGGAACAGAGCG 23032
Qy 61 TGGCTGCGGAGCAGGAGATCGCGCAGCTGATGACTCGCG 101
Db 23033 TATCTGCGCTGCTGGGATCGGCATGTTCGATTCGCG 23073

LOCUS	AX153789	42717 bp	DNA	linear	PAT 22-JUN-2001
DEFINITION	Sequence 113 from Patent WO0140497.				
ACCESSION	AX153789				
VERSION	AX153789.1	GI:14535425			
KEYWORDS					
SOURCE					
ORGANISM					
REFERENCE					
AUTHORS	1 (bases 1 to 42717) Jeannin,P., Pernodet,J.L., Guerinneau,M., Simonet,P., Courtois,S., Cappellano,C., Francon,F., Raynal,A., Ball,M., Sezouov,G., Tuphile,K. and Frostegard,A. Method for obtaining nucleic acids from an environment sample, resulting nucleic acids and use in synthesis of novel compounds Patent: WO 0140497-A 113 07-JUN-2001; Aventis Pharma S.A. (FR)				
TITLE					
JOURNAL					
FEATURES					
source	location/Qualifiers 1..42717 /organism="synthetic construct" /db_xref="taxon:32630" /note="cosmid a2691 brin non codant"				
BASE COUNT	8230 a 13520 c 13184 g 7782 t			1 others	
ORIGIN					
Query Match	52.5%; Score 53; DB 6; Length 42717;				
Best Local Similarity	70.3%; Pred. No. 0.0035;				
Matches	71; Conservative 0; Mismatches 30; Indels 0; Gaps 0;				
OY	1 ATCGCCGCCACCTTGGGGCGGAGATATTGTGGACCCCTGTGACCCGCGAGACGGGCG 60				
Db	11147 ATCGCAGACGCTGCGCGCGGAGATCTTGTCTACTGCGGAGATCCGAAAAACGACGC 11088				
OY	61 TGGCTCGCGAGACAGGATCGCGACGCGATGATGACTCGCG 101				
Db	11087 TATCTGCGCTCGCTGGCATCGCGCATGTTCGATTCGCG 11047				
RESULT 13					
LOCUS	AX069081	6333 bp	DNA	linear	PAT 25-JAN-2001
DEFINITION	Sequence 61 from Patent WO0102555.				
ACCESSION	AX069081				
VERSION	AX069081.1	GI:12578964			
KEYWORDS					
SOURCE					
ORGANISM					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
FEATURES					
source	location/Qualifiers 1..6333 /organism="Mycobacterium tuberculosis" /db_xref="taxon:1773" /note="mas gene, mycocerosic acid synthase."				
BASE COUNT	1010 a 2010 c 2202 g 1111 t				
ORIGIN					
Query Match	50.9%; Score 51.4; DB 6; Length 6333;				
Best Local Similarity	69.3%; Pred. No. 0.01;				
Matches	70; Conservative 0; Mismatches 31; Indels 0; Gaps 0;				
OY	1 ATCGCCGCCACCTTGGGGCGGAGATATTGTGGACCGCTGTGACCGCGAGAACGGGCG 60				

Db	4732	ATTCGCCCGCCCAAGGAGCGAGAGTTTTCGCCGACCGCGGCAATCCCGCAAGGAGACC	4791
QY	61	TGGCTGCGCGAGCAGAGGATCGCGACAGCTGATGACTCGCG	101
Db	4792	ATGCTGCGCAGACATGGCGCTGACGATGTCTACGATTCGCG	4832
RESULT 14			
MSGMYACSYN			
LOCUS		9699 bp DNA linear	BCF 26-Apr-1993
DEFINITION		Mycobacterium bovis mycoceroid acid synthase gene, complete cds.	
ACCESSION			
VERSION		M58808.1 GI:149977	
KEYWORDS		mycoceroid acid synthase.	
SOURCE		Mycobacterium tuberculosis (strain BCG, pathovar bovis) DNA.	
ORGANISM		Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium tuberculosis complex.	
REFERENCE		1 (bases 1 to 9699)	
AUTHORS		Mathur, M. and Kolatukudy, P. E.	
TITLE		Molecular cloning and sequencing of the gene for mycoceroid acid synthase, a novel fatty acid elongating multifunctional enzyme, from Mycobacterium tuberculosis var. bovis Bacillus Calmette-Guerin J. Biol. Chem. 267 (27), 19388-19395 (1992)	
JOURNAL			
MEDLINE		92406887	
PUBMED		1527058	
FEATURES			
source		location/Qualifiers	
		1..9699	
		/organism="Mycobacterium tuberculosis"	
		/strain="BCG"	
		/db_xref="taxon:1773"	
		complement(153..1049)	
		/note="ORF"	
		/codon_start=1	
		/transl_table=1	
		/protein_id="AA25367.1"	
		/db_xref="GI:149979"	
		/translation="MLHDVVHGRSCSENGRRRTTQYRIGFTIGNALMNRKRIGDAA GLORAEKQDLVLESTRNNAAGRETTGLTLPATIGYVAVYAGVCSATLASHLAS IHRREGVMDLTPTRRNAAGRETTGLTLPATIGYVAVYAGVCSATLASHLAS SVAGCRILAKGPPSDTGVCDVAPPRGRRLTOLRCARHAAGTENNCGATVLRKNSP MTAPNWLALPSVELRMSNSPATSPRLPEOLHAAPATWGLHPRSEVRLGRHMCSC" 1741..8073	
		/function="fatty acid elongating enzyme"	
		/codon_start=1	
		/evidence=experimental	
		/transl_table=1	
		/product="mycoceroid acid synthase"	
		/protein_id="AA25369.1"	
		/db_xref="GI:149980"	
		/translation="MESRVMPPVAVTGMGRPLPGGINSPKIMESILKRGDILLETTPD RMADAYDPEPGVGRSVRMGCGFLDDVAGFDPAEFGISERATSDPOOLMIESR WEATHEAGLDPASLAGSIAVFTGLTHEDVLVLTITAGIASLPVYVIGLNNSVAGRT AHTLGLHPAMTPEDTACSSGLMAVHLACRLHGERDLALAGCAVLEBACAAASA OGMSISGRCHSPEDADGDFRSGECMLVILKRLDILRGNGRTFAVIRGTATODOR TETLITSEDAOVAAYVRAALAAAGVOPEYVGAAGHGTGPIGDIYRSILARYGAG TPLALGSAKMSGSTASGTGVIKRLSLRGGVYPLIHPNLPDELSDVENGLEFV PQAVTPPNKNDHTPRKRVAVSSFGMSNTVHAILEENPAASAPDESPGABVGRPLF MLSTSSDADALQTRQRLATWVEEHODCVASADLAVTLARGRHPRVTAVAANLPEL VEGIRVADADALDYDAVVGHDGRPVVFGGGSQWMAAMGTOLLASLPVAAITAKLE	

FVIAASEGSEVTEALTAQTOYVJDKVQVACVAVYALALTAEMKOTYVGRGAVYGDHDM
 GSAEAAYVAGSLEDAVPAVITCRSTKMTARIVAGAGMSYELPAKOTYNSLMAKGTID
 VVSVYVSPQSTVICTGSDTVBDLIRMEBQVDMANEAVADVASHSQVDPLIDDLAAV
 ALADIAPIPTKPVYSATLEDEDRQPCVCGAWVMD,RNTVQPAAVVQAMQEGYSAF
 ADVSHAPLTHIAPEQOTGSDLSMSVALAMAREBQPLPHIGLITGLTEHRAAGALDYS
 YAPGRALVDAPI,PAMTHARL,FIDDDQGEORAGACITVGHILSHYVRLFEPEERHVA
 QCGDTSVLSLSDHQHNVVALPCAYCEMALAAVYRGAABYRDTFEBOMLLD
 EOPTIDAYASIDAPEVNTVETNRQGETTHRAVALAAAEODCCPPQDITBOLLOAH
 PHAVNGTLMRESFAARVYLGALFGLITVTRRPRRWRKPCSPRDSGSAAPHE
 STRCRMTLVSSSAAASSTYGGWMPVAVGACOPACLRAPPOCPILLHADVQJORD
 PRMEDLIDVGHGVYDLVRLMGTGSENERBEPVISELRLTGMOORALPEYD
 GEAGSWLIDTNSNVDTPMLASTLIDAKSHSPQSTECASLSMSYQDTPNDQACE
 KLSQJLRQRDGVYIYGPVPGQDDEHSLAGRKQVHVRVITELAEFEELRPLFVY
 TROADIVPRHSGEAMNLAQALRGELRYTSHSHPMLRTLLIDVDEHTDVERAOOLL
 GSEDETEPMARKQDQVARYLTSPGLHEERRVAVLDPDHGMSVORPPDLOLLEVY
 ASDRPPRGQTEVAVMSINSADVALFAGRPRLIDDBREPOLGMSFVGYVAVGEVY
 TGHQVGDVRYGSEGGCWRTEFLTCOANLAVTEPRGLTDEQATIAATTAHAATVAGLNDL
 AOIKAGQDVILHSAVGGQQAALSTARAGAVTEFAVANGNPAKMAKMDQVEHYVQSR
 EYFEEQDQJTRPDRDQGVDLIDVQSLGAAQARAGLLETFEGREVEIKADVYQNTRLGI
 PFPRGLFVEYDLALMSYQDNRVGLLEAVYAKELFADQGLVAPDCHVRLALAEADA
 IAAMSAHEHTKQVILVPSGRBSVAVTPQRPLRYRQDGSYITGGJAGGLGLFPAASKIA
 AAGSCRIVLTPARSOPNPKARQTEGIRAGAGADIVCGCINAEEDTADVRLSAAVATGL
 PLRGVLSAAVAVEDATLNTIDTELIDRMSPVYFGSMMHRAVLRADQPLDMFCSSGA
 ALLGSPGQAAVAANSVWDYFHHMPACQALPSAIIAMGAMEGVATRLAEGEIMTE
 PEDEGVAEETIVRHRASVGYIPIAPLADLVRSPPMGEMASVNGOBSRSGKPEM
 ELISLPOBDMAGRERLRLLVEQASVILRRTIDADRSTIEVGLDLSGLHLEMTHTHETEG
 RLRLPVLATNTATARALAOYLDTLAEBOAAPAAS"
 complement(8468, .9532)

BASE COUNT	1691 a	2977 c	3280 g	1751 t
ORIGIN				

Query Match	50.9%	Score 51.4	DB 1	Length 9659
Best Local Similarity	69.3%	Pred. No. 0.0097		
Matches 70; Conservative	0	Mismatches 31	Indels 0	Gaps 0

QY 1 ATCCCGCCGACCTCGCGGGGAGGAAATTATTCGACCCGCTGTATACGGCGGGAAGGGGGG 60
Db 6469 ATTCCCGCCGCAAGGAGGAGGAGATTTTCCGACCCCGCGCATTCGCCGGAAGGAGCC 6520
QY 61 TGCTGCGCGAGCAGGGGATCGCGCACGTATGACTCGG 101
Db 6529 ATGCTGGCGCATGGGCGTGCAGCATGTCTACGAAATTCGG 6569

AE007123/C	AE007123	14326 bp	DNA	linear	BCT 27-APR-2001
LOCUS					
DEFINITION	Mycobacterium tuberculosis CDC1551,	section 209 of 280 of the			
ACCESSION	complete genome.				
VERSION	AE007123 AE000516				
KEYWORDS	AE007123.1 GI:13882792				
SOURCE					
ORGANISM	Mycobacterium tuberculosis CDC1551. Mycobacterium tuberculosis CDC1551. Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium tuberculosis complex. 1 (bases 1 to 14326)				
REFERENCE	Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O., Peterson, J., Deboy, R., Dodson, R., Gwinn, M., Hatt, D., Hickey, E., Kolony, J.F., Nelson, W.C., Umayam, L.A., Ernolaeva, M.,				
AUTHORS					

TITLE Whole genome comparison of *Mycobacterium tuberculosis* clinical and laboratory strains

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 14326)

AUTHORS Flaischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O., Peterson, J., Deboy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E., Kolonay, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M., Salzberg, S.L., Delcher, A., Uterback, T., Weidman, J., Khouri, H., Gill, J., Mikula, A. and Bishai, W.

TITLE Direct Submission

JOURNAL Submitted (23-APR-2001) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA

FEATURES Location/Qualifiers

source 1. .14326

gene	7067. .8809	/gene="WT3011"
CDS	7067. .8808 <td>/gene="WT3011"</td>	/gene="WT3011"
	/note="similar to PID:1658531: identified by sequence"	

```
similarity: putative"
/codon_start=1
/transl_table=11
/product="acyl-CoA synthase"
/protein_id="AAK4738.1"
/db_xref="GI:13882794"
/translation="MSVRSILPALRACARLOPHDPATFMDYEDQMDGVAITLWSQL
YRRLTNAQELSCSTGDRVVISAPQGLEVVAVFALQAGRIAVPLSPVGGVYTD
RSDSVLSDSSPAVLTITSSAVDYVQVAVRARGSPSPSITFVDLLDLDAPNGYFKED
EYSPATILQYTSSTRTTPAGVYMSHONRVFEOJMSYFADTDCIPPNKALSWLP
FYHDMGLVIGCAPILGSPAVLTPSVSLORPARMMHMLASDFHAFSAENFAFELA
AARVTTDDMAGRDGNILITLSGERVOAATIKRFADRFARFNLOERYIRSYGLAEA
TVVATSKPGQPEPVDFTESLSAGHAPCAGGATSLSIMLPRESYIVIVSDTC
IECPDGVGEIWMGVNDVANGYWMQKPDSESTFGKATVTPSPGPEGNLRTGSGVF
TDCMFIITGRKIDLIVYGRNHSPDIEATIQETTRGCAISVPGDSSTKLVATTE
LKRKGDSDQDAMARLGAIKREVTSALSSHGLSADVLVAGSIPITSSGVRGAC
VEQYRQDQFARLDA"
/gene="WT3012"
/802..11564
/802..11564
/gene="WT3012"
/note="similar to PID:699243 SP:P54881; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="membrane protein, Mmpl family"
/protein_id="AAK4739.1"
/db_xref="GI:13882795"
/translation="MSPAGRLHRIYIRLKSSPDCRATITSGSADQQRSPRLTNL
LVAAVAAAVIANTLLFTQAEPHDTSFALLPDQAKTAATSRIAQFPOTGSNAIA
YLVGEGSTLEPOQPYDAVAGLADRTRHGSLVLDWMSDPVYAPLCITSPDGSATA
MWMRLGEGSTLEPOQAEISDAVRSVLROLPSHGLNASTVPAITNDMPQIATMSQATI
VTVAIVAVIITLLRARLSVRAAIVITLADSLAVAMPAAVGVGHMDGSPVSKML
AAVFTIGTITATATMLAARLGSDAHSAAPTRDSIPAFALGCAYITGTLARTP
ALHGVTAGLGEFVALAASLTIVLPALIALAASSQLPAPTTGAGWTGSLIPVSASA
LGTAAVLAICMLPTIKRMVGAENPTROGGQOVLPGNALPVVVYIKSRDSDRPALIL
AINOVSHRIVEPGRKVESAAAPAGVWPTGQASVLSAAGRIADLQOAGSGFEVATA
IKSMKSIIEOMSGAVDOLDSTVNTLAGRAOQVLDPMFLAARNLKKTETMSLYE
TITHVIGFTNCDDVILCTAMRKVIEPDIYVTOYNNELSTADRSASISOTQMSLSS
APRVAQMRSLAAYRSFVPKLETTIDPAMQIIOASAMLNLSADPDTGEGHHS
RKDLADBSYRIVRESMFSDDGTATRLFLYSQGLDLAAASAAQOLEIAAGAMVYGL
VDSQVTVGGAQOIAAAVVDALIHDAVLAVLITLVVALLAASMRGVRHRAAGVGLAS
YLAALGVSIALMOHLDBRELNALVIVFAVLASGVYLVAGIKAGIADAEATGARS
KGAVSGGAVAPFLAALGGVFGAGLVIVSGGSFVLSQIGTVVVLGLVLIIVQANMLP
TTPGR"
/complement(11518..11733)
/gene="WT3013"
/complement(11518..11733)
/gene="WT3013"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK47340.1"
/db_xref="GI:13882796"
/translation="MGSVEPPIORPRTLPRATARRPLTRPTGNGPYWCLRWYIV
GQPSVAMGLEBAVNAALAMSEATLAAP"
/complement(11758..11928)
/gene="WT3014"
/complement(11758..11928)
/gene="WT3014"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK47341.1"
/db_xref="GI:13882797"
/translation="MTVAGDGDGLVLGSAMCDVASFWGAVAQRELRARENRRISPSV
HARRFNFTNST"
/gene="WT3015"
/12196..13437
/gene="WT3015"
/12196..13437
/gene="WT3015"
/12196..13437
/gene="WT3015"
```

```
/note="similar to GB:AL123456; identified by sequence
similarity: putative"
/codon_start=1
/transl_table=11
/product="IS153, OrfA"
/protein_id="AAK47342.1"
/db_xref="GI:13882798"
/translation="MLTVEDEAIEIRLHRAGLPIKMIARVLIGISKNTVKSALSNQO
PKYERAPGSIYDAVERPRLRLQAYPTMPATVIAERIGMERSTIRVLSARVAELRPY
LPDPASRTTVAGIEACDWFPEITLPGFGQRTAKQLPLVMCAISRWLLAML
LPSRCAEDLFEGWMLLEALGAVPVLWDEGAIGRWGGRSELTECOQFRTGLA
KVLICRPADPEPAKGLIERADHLYERSFLPGHVFASPADFNAGLALVNTRRRL
GCAPDRIGADRAMLSLPVPAPATGCTSLRLPRDHYRCDSDNYSVHPGVIQHRVL
VRADIERHVECDDELVDHDIRMAHQTVSDPAHVEAKVLRHRHFSAAAPVEPQV
OYRSLSDVDAIGVDIDGVA"
/13522..14238
/gene="WT3016"
/13522..14238
/gene="WT3016"
/note="similar to GB:AL123456; identified by sequence
similarity: putative"
/codon_start=1
/transl_table=11
/product="IS153, OrfB"
/protein_id="AAK47343.1"
/db_xref="GI:13882799"
/translation="MSQCRGPIAPAPRTGATKNTMPACSGKCGOPGPMVYRAASAP
HASRGSRWKSSITSLMLVANSATPSHINAPITSSDPATITCFMAPPGIGKTHLVGLA
IRACQAGRIVLEATAAEVVARLAEHNGRIYAEITRLCRYPLLVVEVQIIPPEPA
ANLFQOLVSSRYERASLIYVTSNKAFFRWGEVFGDDVYAAAMIDRLVHAEVVALKGD
SYRLKRDLDGRVPPAGTPEE"
BASE COUNT 2504 a 4772 c 4554 g 2496 t
ORIGIN
Query Match 50.9%; Score 51.4; DB 1; Length 14326;
Best Local Similarity 69.3%; Pred. No. 0.0093;
Matches 70; Conservative 0; Mismatches 31; Indels 0; Gaps 0;
```

Search completed: November 6, 2002, 13:27:29
Job time : 601.625 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 6, 2002, 11:51:54 : Search time 1197.62 Seconds
(without alignments)
1365.823 Million cell updates/sec

Title: US-09-724-876-2_COPY_34350_34450

Perfect score: 101

Sequence: 1 gccctcgcatgatacagcc.....agccgggtgcggcgccacgg 101

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

EST: *
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hlc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rtd:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Match Length	ID	Description
1	33.8	33.5	448	14	D48878
2	32.4	32.1	1108	17	CNS0202E
3	32.2	31.9	688	12	BF292924
4	31.6	31.3	275	13	B1182335
5	31.6	31.3	640	14	B0841536
6	31.4	31.1	453	13	B0532897

C	7	31.4	31.1	489	13	B1479591
C	8	31.4	31.1	596	12	B0839382
C	9	31.4	31.1	1040	14	BG786331
C	10	31	30.7	471	12	BF482815
C	11	31	30.7	479	12	BM586452
C	12	31	30.7	575	12	BF482727
C	13	31	30.7	655	13	BM597575
C	14	31	30.7	681	13	BM589723
C	15	31	30.7	681	13	BM589723
C	16	31	30.7	697	13	BM613694
C	17	31	30.7	699	13	BM602212
C	18	30.8	30.5	711	13	BM635804
C	19	30.4	30.1	483	13	B1779236
C	20	30.4	30.1	285	17	A2578025
C	21	30.4	30.1	340	12	BG609200
C	22	30.4	30.1	404	10	BE234713
C	23	30.4	30.1	427	14	R03023
C	24	30.4	30.1	465	14	BQ168921
C	25	30.4	30.1	523	13	B1338701
C	26	30.4	30.1	560	13	B1370238
C	27	30.4	30.1	570	9	B1364597
C	28	30.4	30.1	593	9	AT238769
C	29	30.4	30.1	642	13	AT109974
C	30	30.4	30.1	658	10	B1238327
C	31	30.4	30.1	665	10	B1364549
C	32	30.2	29.9	288	10	BE489919
C	33	30.2	29.9	384	13	BE426565
C	34	30.2	29.9	458	13	B0284149
C	35	30.2	29.9	497	13	B0284149
C	36	30.2	29.9	627	9	BG947398
C	37	30.2	29.9	663	14	AL821899
C	38	30.2	29.9	745	13	AL821899
C	39	30.2	29.9	1168	14	BQ246031
C	40	30	29.7	414	13	BQ946738
C	41	30	29.7	453	9	BM189251
C	42	30	29.7	573	13	AT049168
C	43	30	29.7	751	10	BM132991
C	44	30	29.7	1184	12	BG718226
C	45	29.8	29.5	530	13	BG329213

ALIGNMENTS

RESULT 1	D48878	448 bp	linear	EST 04-APR-2002
LOCUS	R1C515430A	Rice green shoot	Oryza sativa (japonica cultivar-group)	
DEFINITION	CDNA, mRNA sequence.			
ACCESSION	D48878			
VERSION	D48878.1	GI:702587		
KEYWORDS	EST.			
SOURCE	Oryza sativa (japonica cultivar-group).			
ORGANISM	Oryza sativa (japonica cultivar-group)			
REFERENCE	Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta: Spermatophyta: Magnoliopsida: Liliopsida: Poales: Poaceae; Erihartoideae; Oryzaceae; Oryza.			
AUTHORS	1 (bases 1 to 448)			
TITLE	Sasaki, T., Miyao, A. and Yamamoto, K.			
JOURNAL	Rice CDNA from callus 1995			
COMMENT	Unpublished (1995)			
CONTACT	Contact: Takuji Sasaki			
INSTITUTE	National Institute of Agrobiological Resources			
PROGRAM	Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki			
TELEPHONE	305-8602, Japan			
FAX	Tel: 81-298-38-7441			
EMAIL	Fax: 81-298-38-7468			
PROJECT	Email: tsasaki@abrc.affrc.go.jp, URL: http://rpg.dna.affrc.go.jp/PROJECT="RGP"			
FEATURES	Location/Qualifiers			
FEATURES	1..448			
FEATURES	/organism="Oryza sativa (japonica cultivar-group)"			
FEATURES	/cultivar="Nipponbare"			

```
BASE COUNT 115 a 95 c 153 g 79 t 6 others
ORIGIN
```

Query Match	33.5%	Score 33.8	DB 14	Length 448
Best Local Similarity	60.2%	Pred No. 34		
Matches 56, Conservative	0	Mismatches 37	Indels 0	Gaps 0

Oy 9 GGATCATACGCCGGTGTGCGGTAGTGCCACCAGAGAGCAGCACAGGCCGCCTGCA 68
 | | | | | | | | | | | | | | | |
Db 149 GAAGAGACGGCGGCGGTGGCGAAGAGAGAGTGTGGGAGGTGCGCCCCGCGGCGATTCT 208

```

Qy      69  ATTCACGTTCCGAGCCGGGTGCCGGGCCACGG  101
          |  | | | | | | | | | | | | |
Db      209  GGTCCAGAGCGGAGCCGGAGTCGAGCCACCG  241

```

LOCUS CNS0209E 1108 bp DNA linear GSS 14-MAY-2000
DEFINITION Tetradon nigroviridis genome survey sequence PUC-Ori end of clone
 157114 of library G from Tetradon nigroviridis, genomic survey
 sequence.

ACCESSION	AL209003
VERSION	AL209003.1
KEYWORDS	GSS; genome survey sequence
SOURCE	<i>Tetradodon nigroviridis</i> .
ORGANISM	<i>Tetradodon nigroviridis</i>

REFERENCE
AUTHORS
Roest-Crolius, H., Dastlva, C., Bouneau, L., Fisher, C.,
1 (bases 1 to 1108)
Eukaryotes; Metazoa; Chordata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorphi; Tetraodontiformes;
Tetraodontidae; Tetraodon.
1 (bases 1 to 1108)

TITLE Saurin, W. and Weissenbach, J.
Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence

Unpublished
2 (bases 1 to 1108)
Roest-Crollius, H., Jallion, O., Dasilva, C., Fizames, C., Fisher, C.,
Bonneau, E., Billaud, A., Queclier, F., Saurin, W., Bernot, A. and
Weissenbach, J.

TITLE Characterization and repeat analysis of the compact genome of the
JOURNAL freshwater pufferfish *Tetraodon nigroviridis*
REFERENCE Unpublished
AUTHORS 3 (bases 1 to 1108)
Genoscope.

JOURNAL COMMENT Submitted (12-APR-2000) This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at <http://www.genoscope.cns.fr/tetraodon>.

```
FEATURES
  source      Location/Qualifiers
1.  .1108     /organism="Tetrarodon nigroviridis"
          /db_xref="taxon:99883"
```

BASE COUNT	ORIGIN	Genoscope sequence ID : COAG157BE00
262 a	255 c	309 g
		273 t
		9 others

Query Match	32.1%	Score	32.4	DB	17	Length	1108
Best Local Similarity	58.7%	Pred. No.	98				
Matches	54	Conservative	1	Mismatches	37	Indels	0
						Gaps	0

QY 10 GATGATACGCCCGGTGGCGGTGCATGGCGACCAGGAGCCAGCAGCCGCTGCAA 69
||||| | ||||| : ||||| | ||| ||||| ||

Db 731 GATGAGACGCMGAGAGACGAGGGGACACCGCTGTGAGGCTGTCTCGGAGGACGCTCCAC 730

Qy 70 TTCCACGTTGCGAGCGGAGTCCGCGGCGCCACG 101
||||| - ||||| ||||| - - - - -
Db 791 CTCACCTGAGAGAGAGGCTATTAGGACCGTG 822

RESULT 3	LOCUS	DEFINITION
Bf292924/c	688 bp	mrna
Bf292924		linear
WH2216_E01_J0225		EST 17-NOV-2000
		Triticum turgidum L. var. durum (durum wheat)

DEFINITION WHE2166_E01_J02S Triticum turgidum L. var. durum (du whole plant cDNA library Triticum turgidum cDNA clone WHE2166_E01_J02, mRNA sequence.

ACCESSION	BF292924
VERSION	BF292924.1
KEYWORDS	GI:11223988
SOURCE	EST.
ORGANISM	Triticum turgidum. Triticum turgidum

REFERENCE

1 (bases 1 to 688)

Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
; Triticeae; Triticum.

AUTHORS
Akhunov, E., Anderson, O. D., Chao, S., Chih, A., Choi, D. W., Close, T. J., Fenton, R. D., Han, P. S., Hsia, C. C., Kang, Y., Kianian, P., Lazo, G. R., Miller, R. D., Otto, C., Rausch, C. J., Seaton, C. L., Simons, K., Tong, J. C. and Zhang, D.

TITLE
The structure and function of the expressed portion of the wheat genomes - Whole plant cDNA library from *Triticum turgidum* L. var.

JOURNAL
Unpublished (2000)
Contact: Olin Anderson
COMMENT
Department of Anthropology, University of California, Berkeley

US Department of Agriculture, Agriculture Research Service, Pacific
West Area, Western Regional Research Center
800 Buchanan Street, Albany, CA 94710, USA
Tel.: 5105595773
Fax: 5105505010

Fax: 5105595818
 Email: Candern1@pw.usda.gov
 Sequence have been trimmed to remove vector sequence and low
 quality sequence with phred score less than 20
 Seq primer: Stratagene 5K primer.

```

FEATURES
SOURCE
LOCATION/Qualifiers
1. 688
/organism="Triticum turgidum"
/cultivar="Langdon-16"
/db_xref="taxon:4571"
/clone="WHE216_E01_002"
/clone_1b="Triticum turgidum L. var. durum (durum wheat) whole plant cDNA library"

```

```

/tissue_type="All tissues"
/dev_stage="different growth stages"
/lab_host="E. coli SOLR"
/note="Vector: Lambda uni-ZAP XR, excised phagemid;
Site_1: EcoRI; Site_2: XhoI; Plants were grown in a growth

```

chamber at North Dakota State University (Kianlan, Otto, Simons). Tissues collected from seven-day etiolated seedling leaf, stem, root and seed; leaf from plant at fourth leaf stage; spike from pre-anthesis through 20

days after anthesis; flag leaf; leaf and stem tissue from tillers, and root. Total RNA and poly(A) RNA were prepared from each tissue and then pooled, a cDNA library was made, and the cDNA clones were in vivo excised to give

PBLeadsript phagemids in the TJ Close lab (Akhunov, Chin, Choi, Close, Fenton, Kianian, Otto, Simons, Zhang) at the University of California, Riverside. Plasmid DNA preparations and DNA sequencing were performed in the OD

BASE COUNT	98 a	301 c	160 g	129 t
ORIGIN				

Query Match	31.9%	Score 32.2	DB 12	Length 688
Best Local Similarity	63.6%	Pred. No. 98		
Matches 49; Conservative	0	Mismatches 28	Indels 0	Gaps 0

2Y 8 CGGATGATACGCCGGTGGCGGTGCAGGTCATGGCGACCGAGGAGCGACCGCCCTGC 67

	DB	135	GGCGAGCGGAGATCAGC	152	
RESULT 6	BJ532897		453 bp	mRNA	linear EST 09-AUG-2002
LOCUS	BJ532897 MF01SSB cDNA Oryzias latipes cDNA clone MF01SSB015F01 3'				
DEFINITION	mRNA sequence.				
ACCESSION	BJ532897				
VERSION	BJ532897.1 GI:22191709				
KEYWORDS	EST.				
SOURCE	Japanese medaka.				
ORGANISM	Oryzias latipes				
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
Acanthopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;					
Acanthomorphi; Acanthopterygii; Perciformes; Atherinomorpha;					
Belontiiformes; Adrianchthyidae; Oryziinae; Oryzias.					
REFERENCE	1 (bases 1 to 453)				
AUTHORS	Kohara,Y., Shih-n,T., Kimura,T., Narita,T., Jindo,T. and Takeda,H.				
TITLE	Medaka EST project in Takeda's lab				
JOURNAL	Unpublished (2001)				
COMMENT	Contact: Tadasu Shin-I Center For Genetic Resource Information National Institute of Genetics 111 Yata, Mishima, Shizuoka 411-8540, Japan Tel.: 81-559-81-6856 Fax: 81-559-81-6835 Email: tsuhni@genes.nig.ac.jp. Location/Qualifiers				
FEATURES	Source				
	1..453				
	/organism="Oryzias latipes"				
	/strain="Hd-tr"				
	/db_xref="taxon:8090"				
	/clone="MF01SSB015F01"				
	/clone_lib="MF01SSB cDNA"				
	/sex="mixture of female and male"				
	/tissue_type="whole embryo"				
	/dev_stage="segmentation stage 20 - 25"				
BASE COUNT	73 a 135 c 109 g 129 t				7 others
ORIGIN					
Query Match	31.1%; Score 31.4; DB 13; Length 453;				
Best Local Similarity	61.0%; Pred. No. 1.4e+02;				
Matches	50; Conservative	0;	Mismatches	32;	Indels 0; Gaps 0;
DY	19 CCGGTGGCGTCAGTGCATGGCGAGCCGAGCACGCCGCTCAATTCCAGCTT	78			
Db	344 CCCAGCGAGATCGGAACTTGCCGCCAGAAGACCACACGACGCTCTTAGACGTGGCT	403			
DY	79 GCGAGCCGGTCCCGGCCAGC	100			
Db	404 GCGAGCCTCGTCGCCGCCGACG	425			
RESULT 7					
BI479591/c					
LOCUS	BI479591		489 bp	mRNA	linear EST 28-AUG-2001
DEFINITION	WHR3455_F02.K03S wheat pre-anthesis spike cDNA library Trillium aestivum cdna clone WHR3455_F02.K03, mRNA sequence.				
ACCESSION	BI479591				
VERSION	BI479591.1 GI:15322762				
KEYWORDS	EST.				
SOURCE	bread wheat.				
ORGANISM	Trillium aestivum				
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;					
Spermatophytes; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae					
; Triticeae; Triticum.					
1 (bases 1 to 489)					
Anderson,O.D., Chao,S., Choi,D.W., Close,T.J., Fenton,R.D., Han					
,P.S., Hsia,C.C., Kang,Y., Lazo,G.R., Miller,R., Rausch,C.J.,					
Seaton,C.L. and Tong,J.C.					
The structure and function of the expressed portion of the wheat					
genomes - Pre-anthesis spike cDNA library					

JOURNAL
 COMMENT
 Unpublished (2000)
 Contact: Olin Anderson
 US Department of Agriculture, Agriculture Research Service, Pacific
 West Area, Western Regional Research Center
 800 Buchanan Street, Albany, CA 94710, USA
 Tel: 5105595773
 Fax: 5105595818
 Email: oanderson@pw.usda.gov
 Sequence have been trimmed to remove vector sequence and low
 quality sequence with phred score less than 20
 Seq primer: Stratiogene SK primer.

FEATURES
 source
 location/Qualifiers
 1..489
 /organism="Triticum aestivum"
 /cultivar="Chinese Spring"
 /db_xref="taxon:4565"
 /clone="WHE3455_F02_K03"
 /clone.lib="wheat pre-anthesis spike cDNA library"
 /tissue.type="Spike before anthesis"
 /dev_stage="Adult plant"
 /lab_host="E. coli SOLR"
 /note="Vector: Lambda Uni-ZAP XR, excised phagemid;
 site.1: EcoRI; site.2: XhoI; Plants were grown in the
 greenhouse. Whole spike with awns trimmed, white, green
 and yellow anther were collected and total RNA, and
 poly(A) RNA were prepared, a cDNA library was made, and
 the cDNA clones were in vivo excised to give phlucscript
 phagemids in the T7 Close Lab (Choi, Close, Fenton) at
 the University of California, Riverside. Plasmid DNA
 preparations and DNA sequencing were performed in the OD
 Anderson lab (all other authors)."
 BASE COUNT
 ORIGIN
 68 a 199 c 161 g 61 t

Query Match
 Best Local Similarity 31.1%; Score 31.4; DB 13; Length 489;
 Matches 44; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 2 CCTTGGCGAGTATGACCGCGGTGGTGCAGAGTATGCGACGACGAGCAGCAGCC 61
 ||||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 174 CCTTGGCGAGGAGCCGCCGCGTTGAGTGCAGAGCGACGACGAGGCGGACCGT 115
 ||||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 62 GGCTG 66
 || ||
 Db 114 GCATG 110

RESULT 8
 B0839382/c
 LOCUS B0839382 596 bp mRNA linear EST 08-AUG-2002
 DEFINITION WHE4165_E02_J033S wheat GS whole plant cDNA library Triticum
 aestivum CDNA clone WHE4165_E02_J03, mRNA sequence.
 ACCESSION B0839382
 VERSION B0839382.1 GI:22143704
 KEYWORDS EST.
 SOURCE bread wheat.
 ORGANISM Triticum aestivum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidaeae
 ; Triticeae; Triticum.
 1 (bases 1 to 596)
 Anderson,O.D., Akhunov,E., Chao,S., Crossman,C., Deal,K., Dvorak,J.,
 Lazo,G.R., Rausch,C.J., Wilson,C. and Woo,J.
 The structure and function of the expressed portion of the wheat
 genomes - Chinese Spring whole plant cDNA library
 Unpublished (2002)
 Contact: Olin Anderson
 US Department of Agriculture, Agriculture Research Service, Pacific
 West Area, Western Regional Research Center
 800 Buchanan Street, Albany, CA 94710, USA
 Tel: 5105595773
 Fax: 5105595818
 Email: oanderson@pw.usda.gov

JOURNAL
 COMMENT

the University of California, Riverside. Plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (all other authors)."

BASE COUNT 125 a 129 c 121 g 96 t

Query Match

Best Local Similarity 30.7%; Score 31; DB 12; Length 471; Pred. No. 1.8e+02; Matches 49; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

17 CGCGCGTGGCGGTGACGATGCGACGAGAGAGACCGCCCTTCATTCACG 76

23 CGCGCGCCCAAGTCCGCGGTGCGCGACGCTGCATCTCGGCGCTCCCACTTCGCCG 82

77 TTGCGAGCCGGGTGCGCGG 95

83 GCTCGAGCCTCGCGCGCG 101

RESULT 11

BM586452/c 479 bp mRNA linear EST 25-FEB-2002

LOCUS 17000687314815 A Gam ad cDNA.blood1 Anopheles gambiae cDNA clone

DEFINITION 19600449681104 5', mRNA sequence.

ACCESSION BM586452

VERSION BM586452.1 GI:18882313

KEYWORDS EST.

SOURCE African malaria mosquito.

ORGANISM Anopheles gambiae

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Anopheles.

REFERENCE 1 (bases 1 to 479)

AUTHORS Holt,R.A., Llin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L., Charlbat

TITLE R., Collins,F.H., Venter,J.C. and Hoffman,S.L.

JOURNAL Celera Anopheles gambiae EST project

COMMENT Unpublished (2002)

CONTACT: Holt R.A.

CELERA GENOMICS

45 W. Gude Dr., Rockville, MD 20850, USA

Tel: 2404533151

Fax: 2404534580

Email: HoltR@celera.com

Plate: NU010049V9 row: F column: 02

Seq primer: M13 Reverse.

FEATURES

source

1.479

Location/Qualifiers

/organism="Anopheles gambiae"

/strain="RSP-ST (Reduced susc. to permethrin - std. chromosome)"

/db_xref="taxon:7165"

/clone="19600449681104"

/clone_lib="A.Gam.ad.cDNA.blood1"

/dev_stage="Adult"

/lab_host="DHI0b"

/note="Vector: pSPORT1; Site 1: SalI; Site 2: NotI; Whole adult mosquitoes (mixed sex) frozen on liquid nitrogen 24 hours after human blood feeding. cDNA inserts >500 bp cloned directionally into pSPORT 1. Not 1 site is 3'.

Clones available through the Malaria Research and Reference Reagent Resource Center (www.malaria.mr4.org)"

BASE COUNT 80 a 134 c 180 g 85 t

ORIGIN

Query Match 30.7%; Score 31; DB 13; Length 479;

Best Local Similarity 64.8%; Pred. No. 1.9e+02;

Matches 46; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

9 GGATGATACGCGCGGTGACGATGCGACGAGAGAGACCGCCCTTCACG 68

149 GGGCGATCAGCGCGGTGACGATGCGACGAGAGAGACCGCCCTTCACG 90

69 ATTCCACGTTG 79

Db 89 ACTCTCTGTTG 79

RESULT 12

BF482727

LOCUS WHE2301-2304_B07_B07ZS wheat pre-anthesis spike cDNA library

DEFINITION Triticum aestivum cDNA clone WHE2301-2304_B07_B07, mRNA sequence.

ACCESSION BF482727

VERSION BF482727.1 GI:11566028

KEYWORDS EST.

SOURCE bread wheat.

ORGANISM Triticum aestivum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae

; Triticaceae; Triticum.

REFERENCE 1 (bases 1 to 575)

AUTHORS Anderson,O.D., Chao,S., Choi,D.W., Close,T.J., Fenton,R.D., Han

,P.S., Hsia,C.C., Kang,Y., Lazo,G.R., Miller,R., Rausch,C.J.,

Seaton,C.L. and Tong,J.C.

TITLE The structure and function of the expressed portion of the wheat

JOURNAL genomes - Pre-anthesis spike cDNA library

COMMENT Unpublished (2000)

CONTACT: Olin Anderson

US Department of Agriculture, Agriculture Research Service, Pacific

West Area, Western Regional Research Center

800 Buchanan Street, Albany, CA 94710, USA

Tel: 510595773

Fax: 510595818

Email: oanderson@w.usda.gov

Sequence have been trimmed to remove vector sequence and low

quality sequence with phred score less than 20

Seq primer: StrataGene SK primer.

FEATURES

source

1.575

Location/Qualifiers

/organism="Triticum aestivum"

/cultivar="Chinese Spring"

/db_xref="taxon:4565"

/clone="WHE2301-2304_B07_B07"

/clone_lib="Wheat pre-anthesis spike cDNA library"

/tissue_type="Spike before anthesis"

/dev_stage="Adult plant"

/lab_host="E. coli SOLR"

/note="Vector: Lambda Uni-ZAP XR, excised phagemid; Site 1: EcoRI; Site 2: XhoI; Plants were grown in the greenhouse. Whole spike with awns trimmed, white, green and yellow anther were collected and total RNA, and poly(A) RNA were prepared, a cDNA library was made, and the cDNA clones were in vivo excised to give pluscript phagemids in the T7 Close lab (Choi, Close, Fenton) at the University of California, Riverside. Plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (all other authors)."

BASE COUNT 164 a 148 c 152 g 111 t

ORIGIN

Query Match 30.7%; Score 31; DB 12; Length 575;

Best Local Similarity 62.0%; Pred. No. 1.9e+02;

Matches 49; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

17 CGCGGTGCGGTGACGATGCGACGAGAGAGACCGCCCTTCATTCACG 76

23 CGCGCGCCCAAGTCCGCGGTGCGCGACGCTGCATCTCGGCGCTCCCACTTCGCCG 82

77 TTGCGAGCCGGGTGCGCGG 95

83 GCTCGAGCCTCGCGCGCG 101

RESULT 13

BM597575/c 655 bp mRNA linear EST 25-FEB-2002

DEFINITION 1700068749053 A.Gam.ad.cDNA.blood1 Anopheles gambiae cDNA clone
 ACCESSION 1600449683838 5', mRNA sequence.
 VERSION BM597575
 KEYWORDS BM597575.1 GI:18895678
 SOURCE EST.
 ORGANISM African malaria mosquito.
 Anopheles gambiae
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
 Anopheles.
 REFERENCE 1 (bases 1 to 655)
 AUTHORS Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L., Charlab
 'R., Collins,F.H., Venter,J.C. and Hoffman,S.L.
 TITLE Celera Anopheles gambiae EST project
 JOURNAL Unpublished (2002)
 COMMENT Contact: Holt R.A.
 Celera Genomics
 45 W. Gude Dr., Rockville, MD 20850, USA
 Tel: 2404533151
 Fax: 2404534580
 Email: HoltRA@celera.com
 Plate: NU01004825 row: G column: 24
 Seq primer: M13 Reverse.
 Location/Qualifiers
 1..655
 /organism="Anopheles gambiae"
 /strain="RSP-ST (Reduced susc. to Permethrin - std.
 chromosome)"
 /db_xref="taxon:7165"
 /clone="19600449683838"
 /clone_1id="A.Gam.ad.cDNA.blood1"
 /dev_stage="Adult"
 /lab_host="DH10b"
 /note="Vector: pSport1; Site_1: SalI; Site_2: NotI; Whole
 adult mosquitoes (mixed sex) frozen on liquid nitrogen 24
 hours after human blood feeding. cDNA inserts >500 bp
 cloned directionally into pSport 1. Not 1 site is 3'.
 Clones available through the Malaria Research and
 Reference Reagent Resource Center (www.malaria.mr4.org)"
 BASE COUNT 119 a 187 c 233 g 116 t
 ORIGIN
 Query Match 30.7%; Score 31; DB 13; Length 655;
 Best Local Similarity 64.8%; Pred. No. 2e+02;
 Matches 46; Conservative 0; Mismatches 25; Indels 0; Gaps 0;
 QY 9 GGATGATACGCCGGTGGCGGTGACAGTCATGCGACGACGAGGACGACGCCGCTGCA 68
 DB 350 GGGGATGACGCCGGTGGCGGTGACAGTCATGCGACGACGAGGACGACGCCGCTGCA 291
 QY 69 ATTCCACGTTG 79
 DB 290 ACTCCTCGTTG 280
 RESULT 14
 LOCUS BM589723 681 bp mRNA linear EST 25-FEB-2002
 DEFINITION 17000687440309 A.Gam.ad.cDNA.blood1 Anopheles gambiae cDNA clone
 ACCESSION 1600448421008 5', mRNA sequence.
 VERSION BM589723
 KEYWORDS BM589723.1 GI:18885584
 SOURCE EST.
 ORGANISM African malaria mosquito.
 Anopheles gambiae
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
 Anopheles.
 REFERENCE 1 (bases 1 to 681)
 AUTHORS Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L., Charlab
 'R., Collins,F.H., Venter,J.C. and Hoffman,S.L.
 TITLE Celera Anopheles gambiae EST project
 JOURNAL Unpublished (2002)

COMMENT Contact: Holt R.A.
 Celera Genomics
 45 W. Gude Dr., Rockville, MD 20850, USA
 Tel: 2404533151
 Fax: 2404534580
 Email: HoltRA@celera.com
 Plate: NU01004911 row: F column: 07
 Seq primer: M13 Reverse.
 Location/Qualifiers
 1..681
 /organism="Anopheles gambiae"
 /strain="RSP-ST (Reduced susc. to Permethrin - std.
 chromosome)"
 /db_xref="taxon:7165"
 /clone="19600448421008"
 /clone_1id="A.Gam.ad.cDNA.blood1"
 /dev_stage="Adult"
 /lab_host="DH10b"
 /note="Vector: pSport1; Site_1: SalI; Site_2: NotI; Whole
 adult mosquitoes (mixed sex) frozen on liquid nitrogen 24
 hours after human blood feeding. cDNA inserts >500 bp
 cloned directionally into pSport 1. Not 1 site is 3'.
 Clones available through the Malaria Research and
 Reference Reagent Resource Center (www.malaria.mr4.org)"
 BASE COUNT 142 a 198 c 221 g 120 t
 ORIGIN
 Query Match 30.7%; Score 31; DB 13; Length 681;
 Best Local Similarity 64.8%; Pred. No. 2e+02;
 Matches 46; Conservative 0; Mismatches 25; Indels 0; Gaps 0;
 QY 9 GGATGATACGCCGGTGGCGGTGACAGTCATGCGACGACGAGGACGACGCCGCTGCA 68
 DB 116 GGGGATGACGCCGGTGGCGGTGACAGTCATGCGACGACGAGGACGACGCCGCTGCA 57
 QY 69 ATTCCACGTTG 79
 DB 56 ACTCCTCGTTG 46
 RESULT 15
 LOCUS BM613694 687 bp mRNA linear EST 25-FEB-2002
 DEFINITION 17000687139208 A.Gam.ad.cDNA.blood1 Anopheles gambiae cDNA clone
 ACCESSION 1600449702982 5', mRNA sequence.
 VERSION BM613694
 KEYWORDS BM613694.1 GI:18911800
 SOURCE EST.
 ORGANISM African malaria mosquito.
 Anopheles gambiae
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
 Anopheles.
 REFERENCE 1 (bases 1 to 687)
 AUTHORS Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L., Charlab
 'R., Collins,F.H., Venter,J.C. and Hoffman,S.L.
 TITLE Celera Anopheles gambiae EST project
 JOURNAL Unpublished (2002)
 COMMENT Contact: Holt R.A.
 Celera Genomics
 45 W. Gude Dr., Rockville, MD 20850, USA
 Tel: 2404533151
 Fax: 2404534580
 Email: HoltRA@celera.com
 Plate: NU010044AP row: E column: 16
 Seq primer: M13 Reverse.
 Location/Qualifiers
 1..687
 /organism="Anopheles gambiae"
 /strain="RSP-ST (Reduced susc. to Permethrin - std.
 chromosome)"
 /db_xref="taxon:7165"
 /clone="19600449702982"

```

/clone_1lb="A.Gam.ad.cDNA.blood1"
/dev_stage="Adult"
/lab_host="DH10b"
/notes="Vector; site_1: SalI; site_2: NotI; Whole
adult mosquitoes (mixed sex) frozen on liquid nitrogen 24
hours after human blood feeding. cDNA inserts >500 bp
cloned directionally into pSport 1. Not I site is 3'.
Clones available through the Malaria Research and
Reference Reagent Resource Center (www.malaria.mz4.org)"
BASE COUNT      145 a      195 c      224 g      123 t
ORIGIN

```

Query Match	30.7%;	Score 31;	DB 13;	Length 687;
Best Local Similarity	64.8%;	Pred. No. 2e+02;		
Matches 46;	Conservative 0;	Mismatches 25;	Indels 0;	Gaps 0;

0y	9	GGATGATATACGGCGGGGCGGTGTCAAGTCATGCGACCGAGAGGACCAAGCGCGCTGCA	68
	164		
Db	164	GGCGCATACGCGCGGTGGGCGCTTCAACCGGTGCAACGGAGACGGAGATTTCTCGACAGCC	105
0y	69	ATTCCACGTTG	79
	104		
Db	104	ACTCTCTGTTG	94

Search completed: November 6, 2002, 15:52:40
Job time : 1201.62 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 6, 2002, 13:32:49 ; Search time 27.125 Seconds
(without alignments)
1240.503 Million cell updates/sec

Title: US-09-724-876-2_COPY_34350_34450

Perfect score: 101

Sequence: 1 gcttcgcgcatgatacgc.....agccgggtgcgcgcacg 101

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 310279 seqs, 16577418 residues

Total number of hits satisfying chosen parameters: 620558

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications_NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/2/pubpna/PCr_NEM_PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEM_PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubpna/US07_NEM_PUB.seq:*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEM_PUB.seq:*
- 6: /cgn2_6/ptodata/2/pubpna/PCrUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEM_PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubpna/US09_NEM_PUB.seq:*
- 10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/2/pubpna/US10_NEM_PUB.seq:*
- 12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/2/pubpna/US60_NEM_PUB.seq:*
- 14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	30	29.7	167343	10	US-09-962-436-281
2	30	29.7	167343	10	US-09-964-824A-273
3	28.6	28.3	3039	10	US-09-792-127-3
4	28.4	28.1	4127	10	US-09-854-883-10
5	28.4	27.1	324	10	US-09-920-300A-576
6	27.4	27.1	324	10	US-10-033-528-576
7	27.4	27.1	324	10	US-09-799-875-10
8	27.2	26.9	939	10	US-09-021-660A-33
9	27.2	26.9	1622	8	US-08-900-220C-7
10	27.2	26.9	1622	10	US-09-151-999-7
11	27.2	26.9	1985	10	US-09-925-302-332
12	27.2	26.9	2108	10	US-09-954-456-686
13	27.2	26.9	2108	10	US-09-954-456-1117
14	27.2	26.9	2108	10	US-09-880-107-1673
15	27.2	26.9	11220	10	US-09-861-289-32
16	27.2	26.9	36778	10	US-09-861-289-5
17	27.2	26.7	1455	10	US-09-815-242-7924
18	27	26.7	1455	10	US-09-815-242-7924
19	26.8	26.5	2100	10	US-09-948-777-1

c	20	26.6	26.3	298	10	US-09-294-093B-4898	Sequence 4898, Ap
	21	26.6	26.3	603	10	US-09-893-737-267	Sequence 267, App
	22	26.6	26.3	1101	10	US-09-438-623A-3	Sequence 3, App11
	23	26.2	25.9	315	10	US-09-867-701-4939	Sequence 4939, Ap
	24	26.2	25.9	475	10	US-09-864-761-6200	Sequence 6200, Ap
	25	26.2	25.9	510	10	US-09-867-701-3223	Sequence 3223, Ap
	26	26.2	25.9	111282	12	US-10-094-989-3	Sequence 3, App11
	27	26.2	25.7	425	10	US-09-864-761-690	Sequence 690, App
	28	26	25.7	2027	10	US-09-867-550-887	Sequence 887, App
	29	26	25.7	3236	10	US-09-954-456-43	Sequence 43, App1
	30	26	25.7	3936	10	US-09-919-172-49	Sequence 49, App1
	31	25.8	25.5	453	10	US-09-887-576-797	Sequence 797, App
	32	25.8	25.5	847	10	US-09-833-381-44	Sequence 44, App1
	33	25.8	25.5	2223	10	US-09-897-214-13	Sequence 13, App1
	34	25.6	25.3	240	10	US-09-960-352-1820	Sequence 1820, Ap
	35	25.6	25.3	411	10	US-09-960-352-5646	Sequence 5646, Ap
	36	25.6	25.3	419	10	US-09-960-352-8465	Sequence 8465, Ap
	37	25.6	25.3	427	10	US-09-960-352-8465	Sequence 8465, Ap
	38	25.6	25.3	1891	10	US-09-969-708-146	Sequence 146, App
	39	25.4	25.1	413	10	US-09-960-352-6637	Sequence 6637, Ap
	40	25.4	25.1	918	10	US-09-815-242-7705	Sequence 7705, Ap
	41	25.4	25.1	1077	9	US-10-104-340-1	Sequence 1, App11
	42	25.4	25.1	1324	10	US-09-925-301-363	Sequence 363, App
	43	25.4	25.1	1324	10	US-09-925-301-363	Sequence 363, App
	44	25.4	25.1	1521	12	US-10-044-090-56	Sequence 56, App1
	45	25.4	25.1	1788	10	US-09-815-242-8009	Sequence 8009, Ap

ALIGNMENTS

RESULT 1

US-09-962-436-281

Sequence 281, Application US/09962436

Patent No. US20020081301A1

GENERAL INFORMATION:

APPLICANT: Soppet, Daniel

TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Sign

FILE REFERENCE: 689290-75

CURRENT APPLICATION NUMBER: US/09/962,436

PRIOR FILING DATE: 2001-09-25

PRIOR APPLICATION NUMBER: US/60/235,082

PRIOR FILING DATE: 2000-09-25

PRIOR APPLICATION NUMBER: US/60/234,924

PRIOR FILING DATE: 2000-09-25

NUMBER OF SEQ ID NOS: 568

SOFTWARE: PatentIn version 3.0

SEQ ID NO 281

LENGTH: 167343

TYPE: DNA

ORGANISM: Homo sapiens

US-09-962-436-281

Query Match 29.7%; Score 30; DB 10; Length 167343;

Best Local Similarity 67.7%; Pred. No. 1.9;

Matches 42; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY 40 GCGACGAGAGCGACGACGCGCTCCGATTCACGTTGCGAGCGGTCGCGGCGAC 99

Db 117079 GCGTCGAGAGACGCGCGCTCCGAGCTCCGCGGTCGACACCGGTCGAGGCGCG 117138

QY 100 GG 101

Db 117139 GG 117140

RESULT 2

US-09-964-824A-273

Sequence 273, Application US/09964824A

Patent No. US20020102531A1

GENERAL INFORMATION:

APPLICANT: Horrigan, Stephen

```
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Signatu
; FILE OF INVENTION: Sets
; FILE REFERENCE: 689290-73
; CURRENT APPLICATION NUMBER: US/09/964,824A
; CURRENT FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: US/60/236,033
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US/60/236,032
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US/60/236,028
; PRIOR FILING DATE: 2000-09-28
; NUMBER OF SEQ ID NOS: 583
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO: 273
; LENGTH: 167343
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-964-824A-273
```

```
Query Match          29.7%; Score 30; DB 10; Length 167343;
Best Local Similarity 67.7%; Pred. No. 1.9;
Matches 42; Conservative 0; Mismatches 20; Indels 0; Gaps 0;
```

```
OY 40 GCGACCGAGGAGCGACCGCCCTGCAATTCACGTTGCGAGCGGGTCCGGCCAC 99
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 117079 GCGTCGGAAGACCGCGGCGCTCCGACCTCCGCGGTGACACCGGGGTACGGGCGC 117138
```

```
OY 100 GG 101
||
Db 117139 GG 117140
```

```
RESULT 3
US-09-792-127-3/C
; Sequence 3, Application US/09792127
; Patent No. US20020002713A1
; GENERAL INFORMATION:
; APPLICANT: Allen, Steve
; APPLICANT: Beckles, Diane M.
; APPLICANT: Butler, Karla
; APPLICANT: Pearlstein, Rich
; TITLE OF INVENTION: Starch Branching Enzyme IIB
; FILE REFERENCE: BBI439 US NA
; CURRENT APPLICATION NUMBER: US/09/792,127
; CURRENT FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: 60/186098
; PRIOR FILING DATE: 2000-03-01
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: Microsoft Office 97
; SEQ ID NO: 3
; LENGTH: 3039
; TYPE: DNA
; ORGANISM: Trillium aestivum
US-09-792-127-3
```

```
Query Match          28.3%; Score 28.6; DB 10; Length 3039;
Best Local Similarity 55.6%; Pred. No. 3.7;
Matches 55; Conservative 0; Mismatches 44; Indels 0; Gaps 0;
```

```
OY 2 CCTTCGGGATGATAGCCGGTGGCGGTGAGTCAATGGGAGGAGGAGGACGAGGCC 61
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 223 CCGACGCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 164
OY 62 GCGTGAATTCAGCGTTGCGAGCGGGGTGCGCGGCGGCGGCGGCGGCGGCGGCGG 100
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 163 GACTGCAAGCTCAGCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 125
```

```
RESULT 4
US-09-854-883-10
; Sequence 10, Application US/09854883
; Patent No. US20020055479A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Lex M. Cowser
; APPLICANT: Jacqueline Wyalt
; APPLICANT: Susan M. Freier
; APPLICANT: Brett P. Monia
; APPLICANT: Madeline M. Butler
; APPLICANT: Robert McKay
; TITLE OF INVENTION: ANTISENSE MODULATION OF PPIIB EXPRESSION
; FILE REFERENCE: ISPH-0576
; CURRENT APPLICATION NUMBER: US/09/854,883
; CURRENT FILING DATE: 2001-05-14
; PRIOR APPLICATION NUMBER: US 09/629,644
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: US 09/487,368
; PRIOR FILING DATE: 2000-01-18
; NUMBER OF SEQ ID NOS: 389
; SEQ ID NO: 10
; LENGTH: 4127
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (120)...(1418)
US-09-854-883-10
```

```
Query Match          28.1%; Score 28.4; DB 10; Length 4127;
Best Local Similarity 62.9%; Pred. No. 4.4;
Matches 44; Conservative 0; Mismatches 26; Indels 0; Gaps 0;
```

```
OY 18 GCGGTCGCGCGTGCAGTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 77
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 587 GCAGTTGAGTTGAGAGACTGGCTACCGAGGAGGCTCGAGAGATCTGCAATTCACCTA 646
OY 78 TCGAGCCG 87
||| |||
Db 647 CACCACCTGG 656
```

```
RESULT 5
US-09-917-800A-1555
; Sequence 1555, Application US/09917800A
; Patent No. US20020119462A1
; GENERAL INFORMATION:
; APPLICANT: Mendick, Donna
; APPLICANT: Porter, Mark
; APPLICANT: Johnson, Kory
; APPLICANT: Castle, Arthur
; APPLICANT: Elashoff, Michael
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Molecular Toxicology Modeling
; FILE REFERENCE: 44921-5038-US
; CURRENT APPLICATION NUMBER: US/09/917,800A
; CURRENT FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: US 60/222,040
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: US 60/222,880
; PRIOR FILING DATE: 2000-11-02
; PRIOR APPLICATION NUMBER: US 60/290,029
; PRIOR FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: US 60/290,645
; PRIOR FILING DATE: 2001-05-15
; PRIOR APPLICATION NUMBER: US 60/292,336
; PRIOR FILING DATE: 2001-05-22
; PRIOR APPLICATION NUMBER: US 60/295,798
; PRIOR FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: US 60/297,457
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,884
; PRIOR FILING DATE: 2001-06-19
; PRIOR APPLICATION NUMBER: US 60/303,459
; PRIOR FILING DATE: 2001-07-09
; NUMBER OF SEQ ID NOS: 1740
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO: 1555
```


PRIOR APPLICATION NUMBER: 60/049,763
PRIOR FILING DATE: 1997-06-16
NUMBER OF SEQ ID NOS: 42
SOFTWARE: Patentln Ver. 2.1
SEQ ID NO 33
LENGTH: 939
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(939)
US-09-021-660A-33

Query Match 26.9%; Score 27.2; DB 10; Length 939;
Best Local Similarity 61.1%; Pred. No. 8.8;
Matches 44; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

QY 1 GCCTTGGCGATGATACGCCGGTGGCGGTGCAGTCATGCGACCGAGAGCAGCCAGGC 60
DB 445 GACCGGAGAGCCCGACAGGCTGAGAGCCTTCAGTCATGAGACTGAGACCGCCCGCAGC 504
QY 61 CGCCTGCAATTC 72
DB 505 CGCCTGGCACTC 516

RESULT 10

US-08-900-220C-7
Sequence 7, Application US/08900220C
Patent No. US20020045206A1
GENERAL INFORMATION:
APPLICANT: Miao, Ningning
Mang, Monica
Mahanchappa, Nagesh K.
Jin, Ping
Jin, Ping

TITLE OF INVENTION: Method of Treating Dopaminergic and
GABA-nergic Disorders

NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: FOLEY, HOAG & ELIOT LLP
STREET: ONE POST OFFICE SQUARE
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII (text)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/900,220C
FILING DATE: 24-Jul-1997
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Vincent, Matthew P.
REGISTRATION NUMBER: 36,709
REFERENCE/DOCKET NUMBER: ONV-044.01
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 832-1000
TELEFAX: (617) 832-7000

INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:
LENGTH: 1622 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 51..1283
SEQUENCE DESCRIPTION: SEQ ID NO: 7:

US-08-900-220C-7

Query Match 26.9%; Score 27.2; DB 8; Length 1622;
Best Local Similarity 61.1%; Pred. No. 9.2;
Matches 44; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

QY 1 GCCTTGGCGATGATACGCCGGTGGCGGTGCAGTCATGCGACCGAGAGCAGCCAGGC 60
DB 792 GACCGGAGAGCCCGACAGGCTGAGAGCCTTCAGTCATGAGACTGAGACCGCCCGCAGC 851
QY 61 CGCCTGCAATTC 72
DB 852 CGCCTGGCACTC 863

RESULT 11

US-09-151-999-7
Sequence 7, Application US/09151999
Patent No. US20020151460A1
GENERAL INFORMATION:
APPLICANT: Mang, Elizabeth
TITLE OF INVENTION: REGULATION OF EPITHELIAL TISSUE BY HEDGEHOG-LIKE
POLYPEPTIDES, AND FORMULATIONS AND USES RELATED THERETO
FILE REFERENCE: ONV-031.02
CURRENT FILING DATE: 1998-08-11
EARLIER APPLICATION NUMBER: 08/955,552
NUMBER OF SEQ ID NOS: 28
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 7
LENGTH: 1622
TYPE: DNA
ORGANISM: Homo sapien Ihh
FEATURE:
NAME/KEY: CDS
LOCATION: (51)..(1283)
US-09-151-999-7

Query Match 26.9%; Score 27.2; DB 10; Length 1622;
Best Local Similarity 61.1%; Pred. No. 9.2;
Matches 44; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

QY 1 GCCTTGGCGATGATACGCCGGTGGCGGTGCAGTCATGCGACCGAGAGCAGCCAGGC 60
DB 792 GACCGGAGAGCCCGACAGGCTGAGAGCCTTCAGTCATGAGACTGAGACCGCCCGCAGC 851
QY 61 CGCCTGCAATTC 72
DB 852 CGCCTGGCACTC 863

RESULT 12

US-09-925-302-332/C
Sequence 332, Application US/09925302
Patent No. US20020044941A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic acids, proteins and antibodies
FILE REFERENCE: PA104
CURRENT APPLICATION NUMBER: US/09/925,302
CURRENT FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05918
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 896
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 332
LENGTH: 1985
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:


```
; NAME/KEY: misc feature
; LOCATION: (360)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (1985)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-302-332
```

```
Query Match          26.9%; Score 27.2; DB 10; Length 1985;
Best Local Similarity 56.8%; Pred. No. 9.3;
Matches 50; Conservative 0; Mismatches 38; Indels 0; Gaps 0;
```

```
QY 12 TGATACGGCGGTGGCGGTGAGTCATGCGACCGAGGAGCGCGCGCTGCATTT 71
    || || || || || || || || || || || || || || || || || || ||
Db 1482 TGTGAGCTCTCTGGCGGACGAGGTGTAGAGCGGACCGCTCGCTGGAGCCACCA 1423
```

```
QY 72 CCAGTTCGAGCGCGGTGCGCGGCCAC 99
    | | | | | | | | | | | | | | | |
Db 1422 GGAAGGCGCGGCGAGGTGGAAGGCCAC 1395
```

RESULT 13

```
US-09-954-456-686/c
; Sequence 686, Application US/09954456
; Patent No. US20020115057A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Young, Paul
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Canc
; FILE REFERENCE: 689290-76
; CURRENT APPLICATION NUMBER: US/09/954,456
; CURRENT FILING DATE: 2001-09-18
```

```
; PRIOR APPLICATION NUMBER: US/60/233,617
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234,052
; PRIOR FILING DATE: 2000-09-20
```

```
; PRIOR APPLICATION NUMBER: US/60/234,923
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,134
; PRIOR FILING DATE: 2000-09-25
```

```
; PRIOR APPLICATION NUMBER: US/60/235,637
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,638
; PRIOR FILING DATE: 2000-09-26
```

```
; PRIOR APPLICATION NUMBER: US/60/235,711
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,720
; PRIOR FILING DATE: 2000-09-27
```

```
; PRIOR APPLICATION NUMBER: US/60/235,840
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,863
; PRIOR FILING DATE: 2000-09-27
```

```
; SOFTWARE: PatentIn version 3.0
; NUMBER OF SEQ ID NOS: 2276
; SEQ ID NO 686
```

```
; LENGTH: 2108
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-954-456-686
```

```
Query Match          26.9%; Score 27.2; DB 10; Length 2108;
Best Local Similarity 56.8%; Pred. No. 9.3;
Matches 50; Conservative 0; Mismatches 38; Indels 0; Gaps 0;
```

```
QY 12 TGATACGGCGGTGGCGGTGAGTCATGCGACCGAGGAGCGCGCGCTGCATTT 71
    || || || || || || || || || || || || || || || || || || ||
Db 1648 TGTGAGCTCTCTGGCGGACGAGGTGTAGAGCGGACCGCTCGCTGGAGCCACCA 1589
```

```
QY 72 CCAGTTCGAGCGCGGTGCGCGGCCAC 99
    | | | | | | | | | | | | | | | |
Db 1588 GGAAGGCGCGGCGAGGTGGAAGGCCAC 1561
```

```
RESULT 14
US-09-954-456-1117/c
; Sequence 1117, Application US/09954456
; Patent No. US20020115057A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Young, Paul
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using C
; FILE REFERENCE: 689290-76
; CURRENT APPLICATION NUMBER: US/09/954,456
; CURRENT FILING DATE: 2001-09-18
```

```
; PRIOR APPLICATION NUMBER: US/60/233,617
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234,052
; PRIOR FILING DATE: 2000-09-20
```

```
; PRIOR APPLICATION NUMBER: US/60/234,923
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,134
; PRIOR FILING DATE: 2000-09-25
```

```
; PRIOR APPLICATION NUMBER: US/60/235,637
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,638
; PRIOR FILING DATE: 2000-09-26
```

```
; PRIOR APPLICATION NUMBER: US/60/235,711
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,720
; PRIOR FILING DATE: 2000-09-27
```

```
; PRIOR APPLICATION NUMBER: US/60/235,840
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,863
; PRIOR FILING DATE: 2000-09-27
```

```
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1117
; LENGTH: 2108
; TYPE: DNA
```

```
; ORGANISM: Homo sapiens
US-09-954-456-1117
```

```
Query Match          26.9%; Score 27.2; DB 10; Length 2108;
Best Local Similarity 56.8%; Pred. No. 9.3;
Matches 50; Conservative 0; Mismatches 38; Indels 0; Gaps 0;
```

```
QY 12 TGATACGGCGGTGGCGGTGAGTCATGCGACCGAGGAGCGCGCGCTGCATTT 71
    || || || || || || || || || || || || || || || || || || ||
Db 1648 TGTGAGCTCTCTGGCGGACGAGGTGTAGAGCGGACCGCTCGCTGGAGCCACCA 1589
```

```
QY 72 CCAGTTCGAGCGCGGTGCGCGGCCAC 99
    | | | | | | | | | | | | | | | |
Db 1588 GGAAGGCGCGGCGAGGTGGAAGGCCAC 1561
```

```
RESULT 15
US-09-880-107-1673/c
; Sequence 1673, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
```

```
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
```

```
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
```

```
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1673
; LENGTH: 2108
```

```

; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 D50914
US-09-880-107-1673

```

```

Query Match      26.9%; Score 27.2; DB 10; Length 2108;
Best Local Similarity 56.8%; Pred. No. 9.3;
Matches 50; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

```

```

QY 12 TGATAGCCCGGTGGCGGTGTCAGTTCATGGCGACGAGAGCGACGAGCCCGCTGCATT 71
    |||||

```

```

DB 1648 TGGTAGACTCTCTGGCGCAGCAGGTGTAGAGCGGAGCGCTGCGTGGGACGCCACCAACA 1589
    |||||

```

```

QY 72 CCAGTTGCGAGCCGGGTGCGGCCAC 99
    |||||

```

```

DB 1588 GGAAGGGCCGGCGCAGGTGGAAGCCAC 1561
    |||||

```

```

Search completed: November 6, 2002, 20:32:45
Job time : 103.125 secs

```

US-09-335-409-1
Sequence 1 Application MS/09335409

```
; Patent No. 6121029
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/335,409
; PRIORITY FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
; US-09-335-409-1
```

```
Query Match          98.4%; Score 99.4; DB 3; Length 68750;
Best Local Similarity 99.0%; Pred. No. 2.8e-19;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
OY 1 GCCTTGGCGGATGATACGCCGGTGGCGGTGCAGGTGTCATGGCGACCGAGGAGCCAGCAGGC 60
|||||
Db 39962 GCCTTGGCGGATGATACGCCGGCGCGGTGCAGGTGTCATGGCGACCGAGGAGCCAGCAGGC 40021
|||||
```

```
OY 61 CGCCTGCAATTCCAGCTTGGAGCCGGGTGCGCGGCCACCG 101
|||||
```

```
Db 40022 CGCCTGCAATTCCAGCTTGGAGCCGGGTGCGCGGCCACCG 40062
|||||
```

RESULT 3

```
US-09-568-102-1
; Sequence 1, Application US/09568102
; Patent No. 6346404
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,102
; PRIORITY FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIORITY FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
; US-09-568-102-1
```

```
Query Match          98.4%; Score 99.4; DB 4; Length 68750;
Best Local Similarity 99.0%; Pred. No. 2.8e-19;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
OY 1 GCCTTGGCGGATGATACGCCGGTGGCGGTGCAGGTGTCATGGCGACCGAGGAGCCAGCAGGC 60
|||||
Db 39962 GCCTTGGCGGATGATACGCCGGCGCGGTGCAGGTGTCATGGCGACCGAGGAGCCAGCAGGC 40021
|||||
```

```
OY 61 CGCCTGCAATTCCAGCTTGGAGCCGGGTGCGCGGCCACCG 101
|||||
```

```
Db 40022 CGCCTGCAATTCCAGCTTGGAGCCGGGTGCGCGGCCACCG 40062
|||||
```

RESULT 4

```
US-09-567-969-1
```

```
; Sequence 1, Application US/09567969
; Patent No. 6355457
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/567,969
; PRIORITY FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIORITY FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
; US-09-567-969-1
```

```
Query Match          98.4%; Score 99.4; DB 4; Length 68750;
Best Local Similarity 99.0%; Pred. No. 2.8e-19;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
OY 1 GCCTTGGCGGATGATACGCCGGTGGCGGTGCAGGTGTCATGGCGACCGAGGAGCCAGCAGGC 60
|||||
Db 39962 GCCTTGGCGGATGATACGCCGGCGCGGTGCAGGTGTCATGGCGACCGAGGAGCCAGCAGGC 40021
|||||
```

```
OY 61 CGCCTGCAATTCCAGCTTGGAGCCGGGTGCGCGGCCACCG 101
|||||
```

```
Db 40022 CGCCTGCAATTCCAGCTTGGAGCCGGGTGCGCGGCCACCG 40062
|||||
```

RESULT 5

```
US-09-568-480-1
; Sequence 1, Application US/09568480
; Patent No. 6355458
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,480
; PRIORITY FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIORITY FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
; US-09-568-480-1
```

```
Query Match          98.4%; Score 99.4; DB 4; Length 68750;
Best Local Similarity 99.0%; Pred. No. 2.8e-19;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
OY 1 GCCTTGGCGGATGATACGCCGGTGGCGGTGCAGGTGTCATGGCGACCGAGGAGCCAGCAGGC 60
|||||
Db 39962 GCCTTGGCGGATGATACGCCGGCGCGGTGCAGGTGTCATGGCGACCGAGGAGCCAGCAGGC 40021
|||||
```

```
OY 61 CGCCTGCAATTCCAGCTTGGAGCCGGGTGCGCGGCCACCG 101
|||||
```

```
Db 40022 CGCCTGCAATTCCAGCTTGGAGCCGGGTGCGCGGCCACCG 40062
|||||
```

```
RESULT 6
US-09-568-486-1
; Sequence 1, Application US/09568486
; Patent No. 6353459
; GENERAL INFORMATION:
; APPLICANT: Schnupp, Thomas
; APPLICANT: Ligdon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHIIONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,486
; PRIOR FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-486-1

Query Match          98.4%; Score 99.4; DB 4; Length 68750;
Best Local Similarity 99.0%; Pred. No. 2.8e-19;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCCTTCGGGATGATACCGCGGTGGCGTGCAGTCAATGCGAGCAGACGACGAGC 60
Db 39962 GCCTTCGGGATGATACCGCGGTGGCGTGCAGTCAATGCGAGCAGACGAGC 40021
QY 61 GCCTGCAATTCCACGTTGCGAGCCGGGTGCCGGGCCACGG 101
Db 40022 GCCTGCAATTCCACGTTGCGAGCCGGGTGCCGGGCCACGG 40062

RESULT 7
US-09-568-472-1
; Sequence 1, Application US/09568472
; Patent No. 6358719
; GENERAL INFORMATION:
; APPLICANT: Schnupp, Thomas
; APPLICANT: Ligdon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHIIONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,472
; PRIOR FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-472-1

Query Match          98.4%; Score 99.4; DB 4; Length 68750;
Best Local Similarity 99.0%; Pred. No. 2.8e-19;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCCTTCGGGATGATACCGCGGTGGCGTGCAGTCAATGCGAGCAGACGACGAGC 60
Db 39962 GCCTTCGGGATGATACCGCGGTGGCGTGCAGTCAATGCGAGCAGACGAGC 40021
QY 61 GCCTGCAATTCCACGTTGCGAGCCGGGTGCCGGGCCACGG 101
Db 40022 GCCTGCAATTCCACGTTGCGAGCCGGGTGCCGGGCCACGG 40062
```

```
Db 40022 GCCTGCAATTCCACGTTGCGAGCCGGGTGCCGGGCCACGG 40062

RESULT 8
US-09-567-899-1
; Sequence 1, Application US/09567899
; Patent No. 6383787
; GENERAL INFORMATION:
; APPLICANT: Schnupp, Thomas
; APPLICANT: Ligdon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHIIONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/567,899
; PRIOR FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-567-899-1

Query Match          98.4%; Score 99.4; DB 4; Length 68750;
Best Local Similarity 99.0%; Pred. No. 2.8e-19;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCCTTCGGGATGATACCGCGGTGGCGTGCAGTCAATGCGAGCAGACGACGAGC 60
Db 39962 GCCTTCGGGATGATACCGCGGTGGCGTGCAGTCAATGCGAGCAGACGAGC 40021
QY 61 GCCTGCAATTCCACGTTGCGAGCCGGGTGCCGGGCCACGG 101
Db 40022 GCCTGCAATTCCACGTTGCGAGCCGGGTGCCGGGCCACGG 40062

RESULT 9
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; PRIOR FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match          31.3%; Score 31.6; DB 4; Length 4403765;
Best Local Similarity 58.5%; Pred. No. 1.8;
Matches 55; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 1 GCCTTCGGGATGATACCGCGGTGGCGTGCAGTCAATGCGAGCAGACGACGAGCAGC 60
```

|||||
Db 4263132 GCCTGCGGATATATCGAGCGTGCGACAGACAGCCATGACACTTTCGCGC 4263133
QY 61 CGCCTGCAATTCACGTTGGAGCGGGTCCGG 94
Db 4263132 GCGCCGAGACACATCTTTGATCTCGAATCGG 4263099

RESULT 10
US-09-103-840A-1/C
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-2007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 1
; LENGTH: 441529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match 31.3%; Score 31.6; DB 4; Length 441529;
Best Local Similarity 58.5%; Pred. No. 1.8;
Matches 55; Conservative 0; Mismatches 39; Indels 0; Gaps 0;
QY 1 GCCTTGGGATATACGCCGGTGGCGGTGCGATGCGGCGAGCGACCGACCGC 60
Db 4270935 GCCTGCGGATATATCGAGCGTGCGACAGACAGCGCATGACACTTTCGCGC 4270876
QY 61 CGCCTGCAATTCACGTTGGAGCGGGTCCGG 94
Db 4270875 GCGCCGAGACACATCTTTGATCTCGAATCGG 4270842

RESULT 11
US-08-998-416-622/C
; Sequence 622, Application US/08998416
; Patent No. 6239264
; GENERAL INFORMATION:
; APPLICANT: Philippsen, Peter
; APPLICANT: Pohlmann, Rainer
; APPLICANT: Steiner, Sabine
; APPLICANT: Mohr, Christine
; APPLICANT: Wendland, Jurgen
; APPLICANT: Knechtle, Philipp
; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPII
; TITLE OF INVENTION: AND USES THEREOF
; NUMBER OF SEQUENCES: 1152
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6239264artis Corporation
; STREET: 3054 Cornwalls Road
; CITY: Research Triangle Park
; STATE: No. 6239264th Carolina
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/998,416

; FILING DATE: 24-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CH 0016/97
; FILING DATE: 31-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 622:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1152 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: PAG1421UP
US-08-998-416-622

Query Match 29.3%; Score 29.6; DB 4; Length 1152;
Best Local Similarity 61.8%; Pred. No. 4.3;
Matches 47; Conservative 0; Mismatches 29; Indels 0; Gaps 0;
QY 1 GCCTTGGGATATACGCCGGTGGCGGTGCGATGCGGCGAGCGACCGACCGC 60
Db 704 GCGCTGCGACACCGCGCGGTGCGGTCTTGCGAGACGACGACGACGCGCGC 645
QY 61 CGCCTGCAATTCACG 76
Db 644 CTGCTGAAGTCGTCG 629

RESULT 12
US-08-822-586-45
; Sequence 45, Application US/08822586
; Patent No. 6015890
; GENERAL INFORMATION:
; APPLICANT: WILLIAM R. JACOBS, JR., JAMES M. MUSSER AND
; APPLICANT: AMALIO TELENTE
; TITLE OF INVENTION: AN EMBCAB OPERON OF MYCOBACTERIA AND
; TITLE OF INVENTION: MUTANTS THEREOF
; NUMBER OF SEQUENCES: 57
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: AMSTER, ROTHSTEIN & EBENSTEIN
; STREET: 90 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: U.S.A.
; ZIP: 10016
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH 1.44 MB STORAGE
; MEDIUM TYPE: DISKETTE
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/822,586
; FILING DATE: MARCH 20, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: ELIZABETH A. BOGOSIAN
; REGISTRATION NUMBER: 39,911
; REFERENCE/DOCKET NUMBER: 96700/437
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 697-5995
; TELEFAX: (212) 286-0854 or 286-0082
; TELEX: TWX 710-581-4766
; INFORMATION FOR SEQ ID NO: 45:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10095

TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
HYPOTHEICAL: NO
US-08-822-586-45

Query Match 28.7%; Score 29; DB 3; Length 10095;
Best Local Similarity 55.4%; Pred. No. 7;
Matches 56; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 1 GCCTTCGGATGATACCGCGGTGCGGTGAGTATGCGACCGAGACGACGAGC 60
DB 865 GCGCTACACGGCGCGCGCTGTGCTGACATCTGACACCGCGCGCATGCGG 924
QY 61 GCGCTGCAATTCACGTGCGACCGCGGTGCGCGCGCGC 101
DB 925 CACCGCGGTTCCTGCGCGCGCTGTGTCGACCGCGCG 965

RESULT 13
US-09-036-987A-1

; Sequence 1, Application US/09036987A
; Patent No. 6143526
; GENERAL INFORMATION:
; APPLICANT: Baltz, Richard H.
; APPLICANT: Broughton, Mary C.
; APPLICANT: Crawford, Kathryn P.
; APPLICANT: Madduri, Krishnamurthy
; APPLICANT: Merlo, Donald J.
; APPLICANT: Treadway, Patil J.
; APPLICANT: Turner, Jan R.
; APPLICANT: Waldron, Clive
; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
; TITLE OF INVENTION: Production
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dow Agrosciences LLC Patent Department
; STREET: 9330 Zionsville Road
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: USA
; ZIP: 46268
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/036,987A
; FILING DATE: 09-MAR-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Stuart, Donald R
; REGISTRATION NUMBER: 28,479
; REFERENCE/DOCKET NUMBER: 50,608
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (317)337-4816
; TELEFAX: (317)337-4847
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 80161 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-09-036-987A-1

QY 6 CGCGATGATACGCGGTGCGGTGCGGTGAGTATGCGACCGAGAGCGACCGCGCGCT 65
Query Match 28.5%; Score 28.8; DB 3; Length 80161;
Best Local Similarity 58.0%; Pred. No. 8.9;
Matches 51; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

DB 49072 CGTGACCGACACCTGATGCTGCGACGCGCTTGC CGCGCGCGTGCAGAGGAAAC 49131
QY 66 GCAATTCACGCTGCGAGCGCGGTGCGCG 93
DB 49132 GCAGCTGGCACTTCGCGCGCGGTGCGCG 49159

RESULT 14
US-09-370-700-1

; Sequence 1, Application US/09370700
; Patent No. 6274350
; GENERAL INFORMATION:
; APPLICANT: Baltz, Richard H.
; APPLICANT: Broughton, Mary C.
; APPLICANT: Crawford, Kathryn P.
; APPLICANT: Madduri, Krishnamurthy
; APPLICANT: Treadway, Patil J.
; APPLICANT: Turner, Jan R.
; APPLICANT: Waldron, Clive
; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
; FILE REFERENCE: 50489 DIV1
; CURRENT FILING DATE: 1999-08-09
; EARLIER FILING DATE: 1999-08-09
; EARLIER FILING DATE: 1998-03-09
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 80161
; TYPE: DNA
; ORGANISM: Saccharopolyspora spinosa
US-09-370-700-1

Query Match 28.5%; Score 28.8; DB 4; Length 80161;
Best Local Similarity 58.0%; Pred. No. 8.9;
Matches 51; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 6 CGCGATGATACGCGGTGCGGTGCGGTGAGTATGCGACCGAGAGCGACCGCGCGCT 65
DB 49072 CGTGACCGACACCTGATGCTGCGACGCGCTTGC CGCGCGCGTGCAGAGGAAAC 49131
QY 66 GCAATTCACGCTGCGAGCGCGGTGCGCG 93
DB 49132 GCAGCTGGCACTTCGCGCGCGGTGCGCG 49159

RESULT 15
US-09-487-368A-10

; Sequence 10, Application US/09487368A
; Patent No. 6261840
; GENERAL INFORMATION:
; APPLICANT: Lex M. Cowsett
; APPLICANT: Jacqueline Wyatt
; TITLE OF INVENTION: ANTISENSE MODULATION OF PTP1B EXPRESSION
; FILE REFERENCE: RTS-0093
; CURRENT APPLICATION NUMBER: US/09/487,368A
; CURRENT FILING DATE: 2000-01-18
; NUMBER OF SEQ ID NOS: 240
; SEQ ID NO 10
; LENGTH: 4127
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (120)...(1418)
US-09-487-368A-10

QY 18 GCGGTGCGGTGCGGTGCGGTGCGGTGAGTATGCGACCGAGAGCGACCGCGCGCTGCAATTCACGCT 77
Query Match 28.1%; Score 28.4; DB 4; Length 4127;
Best Local Similarity 62.9%; Pred. No. 9.8;
Matches 44; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

Db	587	GCAGTTGGAGTTGGAGAACCTGGCTACCCAGAGGCTCGAGAGATCCTGCATTTCACACTA	646
QY	78	TGGAGCCGG	87
Db	647	CACCACCTGG	656

Search completed: November 6, 2002, 16:56:55
Job time : 1888.62 secs


```
FT FT /tag= d
FT FT /note= "encodes ketide synthase (KS-Y) of the loading
FT FT domain"
FT FT misc_RNA
FT FT 3621..4661
FT FT /tag= e
FT FT /note= "encodes acyl transferase (AT) of the loading
FT FT domain"
FT FT misc_RNA
FT FT 4917..5810
FT FT /tag= f
FT FT /note= "encodes enoyl reductase (ER) of the loading
FT FT domain, potentially involved in formation of the
FT FT thiazole moiety"
FT FT misc_RNA
FT FT 5856..6155
FT FT /tag= g
FT FT /note= "encodes acyl carrier protein (ACP) of the loading
FT FT domain"
FT FT CDS
FT FT 6260..10493
FT FT /tag= h
FT FT /label= epob_gene
FT FT /note= "encodes module 1, the NRPS module"
FT FT misc_RNA
FT FT 2031..3548
FT FT /tag= i
FT FT /note= "encodes condensation domain C2 of the NRPS
FT FT module"
FT FT misc_RNA
FT FT 2031..3548
FT FT /tag= j
FT FT /note= "encodes condensation domain C2 of the NRPS
FT FT module"
FT FT misc_RNA
FT FT 6861..6887
FT FT /tag= k
FT FT /note= "encodes heterocyclization signature sequence"
FT FT misc_RNA
FT FT 6861..6887
FT FT /tag= l
FT FT /note= "encodes condensation domain C4 of the NRPS
FT FT module"
FT FT misc_RNA
FT FT 7358..7366
FT FT /tag= m
FT FT /note= "encodes condensation domain C7 (partial) of the
FT FT NRPS module"
FT FT misc_RNA
FT FT 7898..7921
FT FT /tag= n
FT FT /note= "encodes adenylation domain A1 of the NRPS module"
FT FT misc_RNA
FT FT 7898..7921
FT FT /tag= o
FT FT /note= "encodes adenylation domain A1 of the NRPS module"
FT FT misc_RNA
FT FT 8261..8308
FT FT /tag= p
FT FT /note= "encodes adenylation domain A3 of the NRPS module"
FT FT misc_RNA
FT FT 8411..8422
FT FT /tag= q
FT FT /note= "encodes adenylation domain A4 of the NRPS module"
FT FT misc_RNA
FT FT 8861..8905
FT FT /tag= r
FT FT /note= "encodes adenylation domain A6 of the NRPS module"
FT FT misc_RNA
FT FT 8966..8983
FT FT /tag= s
FT FT /note= "encodes adenylation domain A7 of the NRPS module"
FT FT misc_RNA
FT FT 9090..9179
FT FT /tag= t
FT FT /note= "encodes adenylation domain A8 of the NRPS module"
FT FT misc_RNA
FT FT 9183..9992
FT FT /tag= u
FT FT /note= "encodes oxidation region for forming thiazole"
FT FT misc_RNA
FT FT 10121..10138
FT FT /tag= v
FT FT /note= "encodes adenylation domain A10 of the NRPS
FT FT module"
FT FT misc_RNA
FT FT 10261..10306
FT FT /tag= w
FT FT /note= "encodes thiolation domain (PCP) of the NRPS
FT FT module"
FT FT CDS
FT FT 10639..16137
FT FT /tag= x
```

```
FT FT /label= epoc_gene
FT FT /note= "encodes module 2"
FT FT misc_RNA
FT FT 10654..12033
FT FT /tag= y
FT FT /note= "encodes KS2, the KS domain of module 2"
FT FT misc_RNA
FT FT 12250..13287
FT FT /tag= z
FT FT /note= "encodes AT2, the AT domain of module 2"
FT FT misc_RNA
FT FT 13327..13899
FT FT /tag= aa
FT FT /note= "encodes dehydratase (DH) 2, the DH domain of
FT FT module 2"
FT FT misc_RNA
FT FT 14962..15756
FT FT /tag= ab
FT FT /note= "encodes ketoreductase (KR) 2, the KR domain of
FT FT module 2"
FT FT misc_RNA
FT FT 15763..16008
FT FT /tag= ac
FT FT /note= "encodes ACP2, the ACP domain of module 2"
FT FT CDS
FT FT 16134..37907
FT FT /tag= ad
FT FT /label= epod_gene
FT FT /note= "encodes modules 3-6"
FT FT misc_RNA
FT FT 16425..17606
FT FT /tag= ae
FT FT /note= "encodes KS3"
FT FT misc_RNA
FT FT 17817..18857
FT FT /tag= af
FT FT /note= "encodes AT3"
FT FT misc_RNA
FT FT 19581..20396
FT FT /tag= ag
FT FT /note= "encodes KR3"
FT FT misc_RNA
FT FT 20424..20642
FT FT /tag= ah
FT FT /note= "encodes ACP3"
FT FT misc_RNA
FT FT 20706..22082
FT FT /tag= ai
FT FT /note= "encodes KS4"
FT FT misc_RNA
FT FT 22296..23336
FT FT /tag= aj
FT FT /note= "encodes AT4"
FT FT misc_RNA
FT FT 24069..24647
FT FT /tag= ak
FT FT /note= "encodes KR4"
FT FT misc_RNA
FT FT 24867..25151
FT FT /tag= al
FT FT /note= "encodes ACP4"
FT FT misc_RNA
FT FT 25203..26576
FT FT /tag= am
FT FT /note= "encodes KS5"
FT FT misc_RNA
FT FT 26793..27883
FT FT /tag= an
FT FT /note= "encodes AT5"
FT FT misc_RNA
FT FT 27966..28574
FT FT /tag= ao
FT FT /note= "encodes DH5"
FT FT misc_RNA
FT FT 29433..30287
FT FT /tag= ap
FT FT /note= "encodes ER5"
FT FT misc_RNA
FT FT 30321..30869
FT FT /tag= aq
FT FT /note= "encodes KR5"
FT FT misc_RNA
FT FT 31077..31373
FT FT /tag= ar
FT FT /note= "encodes ACP5"
FT FT misc_RNA
FT FT 31440..32807
FT FT /tag= as
FT FT /note= "encodes KS6"
FT FT misc_RNA
FT FT 33018..34067
FT FT /tag= at
FT FT /note= "encodes AT6"
FT FT misc_RNA
FT FT 34107..34676
FT FT /tag= au
```

```
FT /note= "encodes DH6"
FT misc_RNA 35760..36641
FT /*tag= av
FT /note= "encodes ER6"
FT misc_RNA 36705..37256
FT /*tag= aw
FT /note= "encodes KR6"
FT misc_RNA 37470..37769
FT /*tag= ax
FT /note= "encodes ACP6"
FT CDS 37912..49308
FT /*tag= ay
FT /label= epoE_gene
FT /note= "encodes modules 7 and 8"
FT misc_RNA 38014..39375
FT /*tag= az
FT /note= "encodes KS7"
FT misc_RNA 39589..40626
FT /*tag= ba
FT /note= "encodes AT7"
FT misc_RNA 41341..41922
FT /*tag= db
FT /note= "encodes KR7"
FT misc_RNA 42181..42423

Query Match 100.0%; Score 101; DB 21; Length 71989;
Best Local Similarity 100.0%; Pred. No. 6.3e-20;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCTTCGGGATGATACGCCGCGGTGCAGTGCATGAGCGACGAGCAGCAGG 60
Db 34350 GCCTTCGGGATGATACGCCGCGGTGCAGTGCATGAGCGACGAGCAGCAGG 34409
QY 61 GCCTTCGCAATTCACGTTGCCAGCGCGGTGCCGCGCCACGG 101
Db 34410 GCCTTCGCAATTCACGTTGCCAGCGCGGTGCCGCGCCACGG 34450

RESULT 2
AA255887
ID AA255887 standard; DNA; 68750 BP.
XX
AC AA255887;
XX
DT 10-APR-2000 (first entry)
XX
DE Sorangium cellulosum 68.75 kb contig.
XX
KW Epothilone biosynthesis; type I polyketide synthase; taxol substitute; anticancer; ds.
XX
OS Sorangium cellulosum.
XX
FH Key Location/Qualifiers
FT CDS 1..1826
FT /*tag= a
FT /partial
FT /product= "Partial Orf 1 protein (AAV58580)"
FT /note= "No initiation codon given in the specification"
FT CDS
FT /*tag= b
FT /product= "Orf 2 protein (AAV58581)"
FT 3415..5556
FT /*tag= c
FT /product= "Orf 3 protein (AAV58582)"
FT complement(5612..5992)
FT /*tag= d
FT /product= "Orf 4 protein (AAV58583)"
FT 6226..6675
FT /*tag= e
FT /product= "Orf 5 protein (AAV58584)"
FT 7610..11875
FT /*tag= f
```

```
FT /product= "Type I polyketide synthase, EPOS A
FT (AAV58573)"
FT 11872..116104
FT /*tag= g
FT /product= "Non-ribosomal peptide synthetase, EPOS P
FT (AAV58574)"
FT 16251..21749
FT /*tag= h
FT /product= "Type I polyketide synthase, EPOS B
FT (AAV58575)"
FT 21746..43519
FT /*tag= i
FT /product= "Type I polyketide synthase, EPOS C
FT (AAV58576)"
FT 43524..54920
FT /*tag= j
FT /product= "Type I polyketide synthase, EPOS D
FT (AAV58577)"
FT 54935..62254
FT /*tag= k
FT /product= "Type I polyketide synthase, EPOS E
FT (AAV58578)"
FT 62369..63628
FT /*tag= l
FT /product= "Cytochrome P450 oxygenase homologue, EPOS F
FT (AAV58579)"
FT 63779..64333
FT /*tag= m
FT /product= "Orf 6 protein (AAV58585)"
FT complement(63853..64290)
FT /*tag= n
FT /product= "Orf 7 protein (AAV58586)"
FT 64363..64920
FT /*tag= o
FT /product= "Orf 8 protein (AAV58587)"
FT complement(64287..64727)
FT /*tag= p
FT /product= "Orf 9 protein (AAV58588)"
FT 65063..65767
FT /*tag= q
FT /product= "Orf 10 protein"
FT complement(65008..65874)
FT /*tag= r
FT /product= "Orf 11 protein (AAV58590)"
FT complement(65871..66338)
FT /*tag= s
FT /product= "Orf 12 protein (AAV58591)"
FT 66667..67137
FT /*tag= t
FT /product= "Orf 13 protein (AAV58592)"
FT 67334..68251
FT /*tag= u
FT /product= "Orf 14 protein (AAV58593)"
FT 68346..68750
FT /partial
FT /*tag= v
FT /product= "Partial Orf 15 protein (AAV58594)"
FT /note= "No termination codon given in the specification"
FT

W09966028-A2.
XX
PD 23-DEC-1999.
XX
PF 16-JUN-1999; 99WO-EP04171.
XX
PR 18-JUN-1998; 98US-0099504.
PR 24-SEP-1998; 98US-0101631.
PR 05-FEB-1999; 99US-0118906.
XX
PA (NOVS ) NOVARTIS AG.
PA (NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.
XX Schupp T, Ligon JM, Molnar I, Zirkle R, Goerlach J, Cyr D;
```

```
XX MPI: 2000-097741/08.
DR P-PSDB: AAY58573, AAY58574, AAY58575, AAY58576, AAY58577, AAY58578,
DR AAY58579, AAY58580, AAY58581, AAY58582, AAY58583, AAY58584,
DR AAY58585, AAY58586, AAY58587, AAY58588, AAY58590, AAY58591,
DR AAY58592, AAY58593, AAY58594.
XX
XX New isolated epoethione synthase genes, used for the recombinant
PT production of epoethione for use in cancer therapy -
XX
XX Claim 14: Page 87-104, 174pp: English.
XX
XX This sequence represents a 68.75 kb contig from Sorangium cellulosum
CC comprising 22 open reading frames (ORFs) and includes genes encoding
CC proteins involved in the biosynthesis of epoethiones. Epoethiones A and
CC B are 16-membered macrocyclic polyketides with an acylglycine-derived
CC starter unit; polyketides being synthesised from two-carbon building
CC blocks, the beta-carbon of which always carries a keto group. Each round
CC of two-carbon addition is carried out by a complex of enzymes known as
CC the polyketide synthase in a manner similar to fatty acid biosynthesis.
CC EPOS A (AAY58573) and EPOS P (AAY58574) are involved in formation of
CC the thiazole ring formation of epoethiones, and EPOS B, EPOS C, EPOS D
CC and EPOS E (AAY58575-58578) are involved in polyketide backbone
CC formation. EPO F (AAY58579) is an epoethione macrolactone oxidase, and
CC the proteins Orf 3 (AAY58582) and Orf14 (AAY58593) are thought to be
CC involved in transport. Epoethiones mimic the biological activity of
CC taxol, and may be substituted for taxol in cancer chemotherapeutic
CC compositions. Epoethiones exhibit a much lower drop in potency against a
CC multiply drug-resistant cell line compared with taxol, and are
CC considerably less efficiently exported from such cells by the multidrug
CC resistance protein (MDR, or P-glycoprotein). Despite the potential of
CC epoethiones as anticancer agents, they are problematical to produce on a
CC large scale. Epoethiones are too complex for industrial scale chemical
CC synthesis, and Sorangium cellulosum is difficult to ferment, producing
CC poor yields of epoethiones. The nucleic acids of the invention may be
CC used for the recombinant production of epoethiones in a heterologous host
CC that is more amenable to fermentation.
XX
XX Sequence 68750 BP; 9596 A; 22458 C; 25537 G; 11159 T; 0 other;
SQ
XX
XX Query Match 98.4%; Score 99.4; DB 21; Length 68750;
XX Best Local Similarity 99.0%; Pred. No. 1.8e-19;
XX Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX 1 GCCTTGGCGGATGATACGCCGGTGGCGGTGACAGTATGCGACCGAGCGACCGACGCG 60
DB 39962 GCCTTGGCGGATGATACGCCGGGCGGTGACAGTATGCGACCGAGCGACCGACGCG 40021
XX
XX 61 CGCCTGCAATTCCACGTTGCGAGCCGGGTGCGCGGCCACGG 101
DB 40022 CGCCTGCAATTCCACGTTGCGAGCCGGGTGCGCGGCCACGG 40062
XX
XX RESULT 3
XX ID AAI99683 standard; DNA: 4403765 BP.
XX AC AAI99683;
XX
XX 15-JUN-2002 (first entry)
XX
XX Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 2.
XX
XX Mycobacterium tuberculosis: strain H37Rv; strain CDC 1551; genome;
KM variation; epidemiology; patient treatment; epidemic monitoring; ds.
XX
XX Mycobacterium tuberculosis.
OS
XX US6294328-B1.
XX PN 25-SEP-2001.
XX PD 24-JUN-1998; 98US-0103840.
XX PF
```

```
XX
XX 24-JUN-1998; 98US-0103840.
PR (GENO-) INST GENOMIC RES.
XX
XX Fleischmann RD, White OR, Fraser CM, Venter JC;
PI MPI: 2001-647261/74.
XX
XX Evaluating strain variation of Mycobacterium tuberculosis, comprises
PT determining the nucleotide sequence of the strain at positions in the
PT genome corresponding to positions where M. tuberculosis strains CDC
PT 1551 and H37Rv differ -
XX
XX Claim 4: SEQ ID NO 2; 3pp + Sequence Listing: English.
XX
XX The invention relates to evaluating strain variation within and between
CC different populations of the tuberculosis bacterial pathogen,
CC Mycobacterium tuberculosis or related Mycobacterium by determining the
CC nucleotide sequence of the first strain at positions in the complete
CC sequence of the genome that correspond to positions that differ in the
CC nucleotide sequences of M. tuberculosis strains CDC 1551 (AAI99683) and
CC H37Rv (AAI99682). The method is useful for evaluating strain variation of
CC M. tuberculosis and has valuable application in the fields of
CC tuberculosis genetics, epidemiology, patient treatment and epidemic
CC monitoring.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from USPTO
CC at seqdata.uspto.gov/sequence.html?docid=6294328B1.
XX
XX Sequence 4403765 BP; 757105 A; 1447799 C; 1441301 G; 757371 T; 189 other;
SQ
XX
XX Query Match 31.3%; Score 31.6; DB 22; Length 4403765;
XX Best Local Similarity 58.5%; Pred. No. 8.9;
XX Matches 55; Conservative 0; Mismatches 39; Indels 0; Gaps 0;
XX
XX 1 GCCTTGGCGGATGATACGCCGGTGGCGGTGACAGTATGCGACCGAGCGACCGACGCG 60
DB 4263192 GCCTTGGCGGATGATGCGAGCGGTGCGCATTCAGAACAGCGCATGCGACTTCTCGCC 4263133
XX
XX 61 CGCCTGCAATTCCACGTTGCGAGCCGGGTGCGCGG 94
DB 4263132 GGCCTCGAGCACATCTTTTGATCTCGAATCGG 4263099
XX
XX RESULT 4
XX ID AAI99682 standard; DNA: 4411529 BP.
XX AC AAI99682;
XX
XX 15-JUN-2002 (first entry)
XX
XX Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 1.
XX
XX Mycobacterium tuberculosis: strain H37Rv; strain CDC 1551; genome;
KM variation; epidemiology; patient treatment; epidemic monitoring; ds.
XX
XX Mycobacterium tuberculosis.
OS
XX US6294328-B1.
XX PN 25-SEP-2001.
XX PD 24-JUN-1998; 98US-0103840.
XX PR 24-JUN-1998; 98US-0103840.
XX PN 24-JUN-1998; 98US-0103840.
XX PA (GENO-) INST GENOMIC RES.
XX PI Fleischmann RD, White OR, Fraser CM, Venter JC;
DR MPI: 2001-647261/74.
```

XX Evaluating strain variation of *Mycobacterium tuberculosis*, comprises
PT determining the nucleotide sequence of the strain at positions in the
PT genome corresponding to positions where *M. tuberculosis* strains CDC
PT 1551 and H37Rv differ -
XX
XX Claim 3; SEQ ID NO 1; 3pp + Sequence Listing; English.
PS
CC The invention relates to evaluating strain variation within and between
CC different populations of the tuberculosis bacterial pathogen,
CC *Mycobacterium tuberculosis* or related *Mycobacterium* by determining the
CC nucleotide sequence of the first strain at positions in the complete
CC sequence of the genome that correspond to positions that differ in the
CC nucleotide sequences of *M. tuberculosis* strains CDC 1551 (AA199683) and
CC H37Rv (AA199682). The method is useful for evaluating strain variation of
CC *M. tuberculosis* and has valuable application in the fields of
CC tuberculosis genetics, epidemiology, patient treatment and epidemic
CC monitoring.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from USPTO
CC at seqdata.uspto.gov/sequence.html?docID=6294328B1.
CC
XX Sequence 4411529 BP; 758565 A; 1449983 C; 1444602 G; 758379 T; 0 other;
SQ
Query Match 31.3%; Score 31.6; DB 22; Length 4411529;
Best Local Similarity 58.5%; Pred. No. 8.9;
Matches 55; Conservative 0; Mismatches 39; Indels 0; Gaps 0;
OY 1 GCCTTCGGGATGATACGCCGGTGGCGGTGACATGCGACGAGAGCGACCGAGC 60
Db 4270935 GCCTGCTCGAATAATGTCGACGTGCAATCCAGAACAGCCATCTTCTCGC 4270876
OY 61 GCGCTGCAATTCACGTTGCGAGCGCGGGGCGCGG 94
Db 4270875 GCGCCGAGCACATCTTTGATCTCGAATCGG 4270842
RESULT 5
ABLO6373
ID ABLO6373 standard; cDNA: 947 BP.
XX
AC ABLO6373;
XX
DT 26-MAR-2002 (first entry)
XX
DE *Drosophila melanogaster* expressed polynucleotide SEQ ID NO 13601.
XX
XX *Drosophila*; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
OS *Drosophila melanogaster*.
XX
PN WO200171042-A2.
XX
XX 27-SEP-2001.
XX
PD 23-MAR-2001; 2001WO-US09231.
XX
PE 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE) PE CORP NY.
PA
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI: 2001-656860/75.
DR P-PSDB; ABB62270.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from *Drosophila* and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 13601; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from *Drosophila*. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABLO1840-ABLO30511), expressed DNA
CC sequences (ABBS7737-ABBS72072).
CC (ABBS7737-ABBS72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
CC
SQ Sequence 947 BP; 232 A; 265 C; 291 G; 159 T; 0 other;
SQ
Query Match 30.1%; Score 30.4; DB 23; Length 947;
Best Local Similarity 61.2%; Pred. No. 9.8;
Matches 49; Conservative 0; Mismatches 31; Indels 0; Gaps 0;
OY 7 GCGGATGATACGCCGGTGGCGGTGACATGCGACGAGAGCGACCGCGCTG 66
Db 357 GCCGAGGAAACTGCGCGAGGTGCACTGACAGCTGATGAGAGAGCCCGGACAGGCTG 416
OY 67 CAATTCACAGTTGCGAGCCG 86
Db 417 CACGTCTCGATGTGGCGAG 436
RESULT 6
ABLO6348
ID ABLO6348 standard; cDNA: 2890 BP.
XX
AC ABLO6348;
XX
DT 26-MAR-2002 (first entry)
XX
DE *Drosophila melanogaster* expressed polynucleotide SEQ ID NO 13526.
XX
XX *Drosophila*; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
OS *Drosophila melanogaster*.
XX
PN WO200171042-A2.
XX
XX 27-SEP-2001.
XX
PD 23-MAR-2001; 2001WO-US09231.
XX
PE 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE) PE CORP NY.
PA
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI: 2001-656860/75.
DR P-PSDB; ABB62245.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from *Drosophila* and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 13526; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from *Drosophila*. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABLO1840-ABLO30511), expressed DNA
CC sequences (ABBS7737-ABBS72072).
CC (ABBS7737-ABBS72072).

CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 2890 BP; 784 A; 688 C; 704 G; 714 T; 0 other;

Query Match 30.1%; Score 30.4; DB 23; Length 2890;
Best Local Similarity 61.2%; Pred. No. 11;
Matches 49; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 7 GCGGATGATACGCCGCTGCGGTGCAGTCATGCGACGAGGAGCCGCCCTG 66

DB 2484 GCCGAGGGAACCTGGCGGCTGCACTGATGAGAGAGCCGCCGACAGGCTG 2543

QY 67 CAATTCACGCTTCGAGCCG 86

DB 2544 CACGTCCTGATGTGGGCAG 2563

RESULT 7
ABL06372
ID ABL06372 standard; cDNA; 2947 BP.

XX ABL06372;

AC 26-MAR-2002 (first entry)

XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 13598.

DE Drosophila; developmental biology; cell signalling; insecticide;

KM pharmaceutical; gene; ss.

XX Drosophila melanogaster.

OS WO200171042-A2.

PN 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

PF 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

PI WPI; 2001-656860/75.

DR P-PSDB; ABB62269.

XX New isolated nucleic acid detection reagent for detecting 1000 or more

PT genes from Drosophila and for elucidating cell signalling and cell-cell

PT interactions -

XX Claim 1; SEQ ID NO 13598; 21bp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from Drosophila. The invention is

CC useful in developmental biology and in elucidating cell signalling and

CC cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA

CC sequences (AB01840-ABL16175) and the encoded proteins

CC (AB057737-AB072072).

CC The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 2947 BP; 679 A; 726 C; 799 G; 743 T; 0 other;

Query Match 30.1%; Score 30.4; DB 23; Length 2947;
Best Local Similarity 61.2%; Pred. No. 11;
Matches 49; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 7 GCGGATGATACGCCGCTGCGGTGCAGTCATGCGACGAGGAGCCGCCCTG 66

DB 1357 GCCGAGGGAACCTGGCGGCTGCACTGATGAGAGAGCCGCCGACAGGCTG 1416

QY 67 CAATTCACGCTTCGAGCCG 86

DB 1417 CACGTCCTGATGTGGGCAG 1436

RESULT 8
ABL08162
ID ABL08162 standard; cDNA; 4282 BP.

XX ABL08162;

AC 26-MAR-2002 (first entry)

XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 18968.

DE Drosophila; developmental biology; cell signalling; insecticide;

KM pharmaceutical; gene; ss.

XX Drosophila melanogaster.

OS WO200171042-A2.

PN 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

PF 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

PI WPI; 2001-656860/75.

DR P-PSDB; ABB64059.

XX New isolated nucleic acid detection reagent for detecting 1000 or more

PT genes from Drosophila and for elucidating cell signalling and cell-cell

PT interactions -

XX Claim 1; SEQ ID NO 18968; 21bp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from Drosophila. The invention is

CC useful in developmental biology and in elucidating cell signalling and

CC cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA

CC sequences (AB01840-ABL16175) and the encoded proteins

CC (AB057737-AB072072).

CC The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 4282 BP; 966 A; 1112 C; 1156 G; 1048 T; 0 other;

Query Match 30.1%; Score 30.4; DB 23; Length 4282;
Best Local Similarity 61.2%; Pred. No. 11;
Matches 49; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 7 GCGGATGATACGCCGCTGCGGTGCAGTCATGCGACGAGGAGCCGCCCTG 66

DB 270 GCCGAGGGAACCTGGCGGCTGCACTGATGAGAGAGCCGCCGACAGGCTG 329

QY 67 CAATTCACGCTTCGAGCCG 86

DB 330 CACGTCCTGATGTGGGCAG 349

```
RESULT 9
ABL64403
ID ABL64403 standard; DNA: 167343 BP.
XX
XX ABL64403;
AC
XX
XX 15-MAY-2002 (first entry)
DE Stomach cancer related gene sequence SEQ ID NO:2740.
XX
XX Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
KM stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
KM cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
KW gene; ds.
XX
XX Homo sapiens.
OS
XX
XX WO200194629-A2.
PN
XX
XX 13-DEC-2001.
PD
XX
XX 30-MAY-2001; 2001WO-US10838.
PF
XX
XX 05-JUN-2000; 2000US-209473P.
PR 05-JUN-2000; 2000US-209531P.
PR 18-SEP-2000; 2000US-233133P.
PR 18-SEP-2000; 2000US-233617P.
PR 20-SEP-2000; 2000US-234009P.
PR 20-SEP-2000; 2000US-234034P.
PR 20-SEP-2000; 2000US-234052P.
PR 22-SEP-2000; 2000US-234509P.
PR 22-SEP-2000; 2000US-234567P.
PR 25-SEP-2000; 2000US-234923P.
PR 25-SEP-2000; 2000US-234924P.
PR 25-SEP-2000; 2000US-235077P.
PR 25-SEP-2000; 2000US-235082P.
PR 25-SEP-2000; 2000US-235134P.
PR 25-SEP-2000; 2000US-235280P.
PR 26-SEP-2000; 2000US-235637P.
PR 26-SEP-2000; 2000US-235638P.
PR 27-SEP-2000; 2000US-235711P.
PR 27-SEP-2000; 2000US-235720P.
PR 27-SEP-2000; 2000US-235840P.
PR 27-SEP-2000; 2000US-235863P.
PR 28-SEP-2000; 2000US-236028P.
PR 28-SEP-2000; 2000US-236032P.
PR 28-SEP-2000; 2000US-236033P.
PR 28-SEP-2000; 2000US-236034P.
PR 28-SEP-2000; 2000US-236109P.
PR 28-SEP-2000; 2000US-236111P.
PR 29-SEP-2000; 2000US-236842P.
PR 29-SEP-2000; 2000US-236891P.
PR 02-OCT-2000; 2000US-237172P.
PR 02-OCT-2000; 2000US-237173P.
PR 02-OCT-2000; 2000US-237278P.
PR 02-OCT-2000; 2000US-237294P.
PR 02-OCT-2000; 2000US-237295P.
PR 02-OCT-2000; 2000US-237316P.
PR 03-OCT-2000; 2000US-237425P.
PR 03-OCT-2000; 2000US-237598P.
PR 03-OCT-2000; 2000US-237604P.
PR 03-OCT-2000; 2000US-237606P.
PR 03-OCT-2000; 2000US-237608P.
PR 01-NOV-2000; 2000US-244867P.
PR 01-NOV-2000; 2000US-245084P.
XX
XX (AVAL-) AVALON PHARM.
XX
XX Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
PI Soppet DR, Weaver Z;
XX
XX WPI; 2002-188264/24.
```

```
XX
XX Screening for anti-neoplastic agent involves exposing cells to a
PT chemical agent to be tested for anti-neoplastic activity, and
PT determining a change in expression of a gene of a signature gene set -
XX
XX Claim 1; SEQ ID 2740; 44pp; English.
XX
XX The present invention describes a method (M1) for screening for an
CC anti-neoplastic agent. The method involves exposing cells to a chemical
CC agent to be tested for anti-neoplastic activity, determining a change in
CC expression of at least one gene (I) of a signature gene set, where (I)
CC comprises a sequence (S) selected from 8447 sequences (given in ABL61664
CC to ABL70110), or is at least 95% identical to (S), where a change in
CC expression is indicative of anti-neoplastic activity. (I) has cytostatic
CC activity and can be used in gene therapy. M1 can be used for screening
CC an anti-neoplastic agent, and can be used for producing a product which
CC is the data collected with respect to the anti-neoplastic agent as a
CC result of M1, and the data is sufficient to convey the chemical
CC structure and/or properties of the agent. M1 can be used in the
CC treatment of cancer such as colon, breast, stomach, lung, thyroid,
CC oesophageal, ovarian, kidney, prostate or pancreatic cancer,
CC adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,
CC infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine
CC carcinoma, papillary carcinoma and Wilm's tumour.
XX
XX Sequence 167343 BP; 49550 A; 37098 C; 35753 G; 44929 T; 13 other;
SQ
Query Match 29.7%; Score 30; DB 24; Length 167343;
Best Local Similarity 67.7%; Pred. No. 20;
Matches 42; Conservative 0; Mismatches 20; Indels 0; Gaps 0;
XX
XX
XX 40 GCGACCGAGGAGCGACCGCGCTGCATTCACGTTGGAGCGCGGTCGCCGCCAC 99
QY ||| ||||| ||| ||||| ||||| ||| ||| ||| ||||| |||||
Db 117079 GCGTCCGAAGACCGCGGAGCTCCGCGAGCTCCGCGGTACACCGGGTACGGGCGC 117138
QY 100 gg 101
Db 117139 GG 117140
XX
XX
XX RESULT 10
XX ABL67239
XX
XX ABL67239;
AC
XX
XX 15-MAY-2002 (first entry)
DE
XX
XX Thyroid cancer related gene sequence SEQ ID NO:5576.
DE
XX
XX Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
KW cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
KW gene; ds.
XX
XX Homo sapiens.
OS
XX
XX WO200194629-A2.
PN
XX
XX 13-DEC-2001.
PD
XX
XX 30-MAY-2001; 2001WO-US10838.
PF
XX
XX 05-JUN-2000; 2000US-209473P.
PR 05-JUN-2000; 2000US-209531P.
PR 18-SEP-2000; 2000US-233133P.
PR 18-SEP-2000; 2000US-233617P.
PR 20-SEP-2000; 2000US-234009P.
PR 20-SEP-2000; 2000US-234034P.
PR 20-SEP-2000; 2000US-234052P.
PR 22-SEP-2000; 2000US-234509P.
PR 22-SEP-2000; 2000US-234567P.
PR 25-SEP-2000; 2000US-234923P.
```

PR 25-SEP-2000; 2000US-23492AP.
PR 25-SEP-2000; 2000US-235077P.
PR 25-SEP-2000; 2000US-235082P.
PR 25-SEP-2000; 2000US-235134P.
PR 25-SEP-2000; 2000US-235280P.
PR 26-SEP-2000; 2000US-235637P.
PR 26-SEP-2000; 2000US-235638P.
PR 27-SEP-2000; 2000US-235711P.
PR 27-SEP-2000; 2000US-235720P.
PR 27-SEP-2000; 2000US-235840P.
PR 27-SEP-2000; 2000US-235863P.
PR 28-SEP-2000; 2000US-236028P.
PR 28-SEP-2000; 2000US-236032P.
PR 28-SEP-2000; 2000US-236033P.
PR 28-SEP-2000; 2000US-236034P.
PR 28-SEP-2000; 2000US-236119P.
PR 29-SEP-2000; 2000US-236842P.
PR 29-SEP-2000; 2000US-236891P.
PR 02-OCT-2000; 2000US-237172P.
PR 02-OCT-2000; 2000US-237173P.
PR 02-OCT-2000; 2000US-237278P.
PR 02-OCT-2000; 2000US-237294P.
PR 02-OCT-2000; 2000US-237295P.
PR 02-OCT-2000; 2000US-237316P.
PR 03-OCT-2000; 2000US-237425P.
PR 03-OCT-2000; 2000US-237598P.
PR 03-OCT-2000; 2000US-237604P.
PR 03-OCT-2000; 2000US-237606P.
PR 03-OCT-2000; 2000US-237608P.
PR 01-NOV-2000; 2000US-244867P.
PR 01-NOV-2000; 2000US-245084P.
XX
XX
PA (AVAL-) AVALON PHARM.
PI Young PE, Augustus M, Carter KC, Ehner R, Endress G, Horrigan S;
PI Soppet DR, Weaver Z;
XX
XX
DR WPI: 2002-188264/24.
XX
XX
PT Screening for anti-neoplastic agent involves exposing cells to a
PT chemical agent to be tested for anti-neoplastic activity, and
PT determining a change in expression of a gene of a signature gene set -
XX
XX
PS Claim 1; SEQ ID 5576; 44pp; English.

CC The present invention describes a method (M1) for screening for an
CC anti-neoplastic agent. The method involves exposing cells to a chemical
CC agent to be tested for anti-neoplastic activity, determining a change in
CC expression of at least one gene (I1) of a signature gene set, where (I1)
CC comprises a sequence (S) selected from 8447 sequences (given in ABL61664
CC to ABL70110), or is at least 95% identical to (S), where a change in
CC expression is indicative of anti-neoplastic activity. (I1) has cytostatic
CC activity and can be used in gene therapy. M1 can be used for screening
CC an anti-neoplastic agent, and can be used for producing a product which
CC is the data collected with respect to the anti-neoplastic agent as a
CC result of M1, and the data is sufficient to convey the chemical
CC structure and/or properties of the agent. M1 can be used in the
CC treatment of cancer such as colon, breast, stomach, lung, thyroid,
CC oesophageal, ovarian, kidney, prostate or pancreatic cancer,
CC adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,
CC infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine
CC carcinoma, papillary carcinoma and Wilms' tumour.
XX
XX
SQ Sequence 167343 BP; 49550 A; 37098 C; 35753 G; 44929 T; 13 other;

Query Match 29.7%; Score 30; DB 24; Length 167343;
Best Local Similarity 67.7%; Pred. No. 20;
Matches 42; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

OY 40 GCGACGAGGAGGACCGCCCGCAATTCGACGTTCCGCGCGGCGGCGGCAC 99
DB 117079 GCGTCCGAAAGACCGCGAGCCTCCGCGAGCTCCGCGTACACCGCGGTACAGGCGCGC 117138

OY 100 GC 101
DB 117139 GC 117140

RESULT 11
AA253393/c
ID AA253393 standard; DNA; 1620 BP.

XX AA253393;

DT 21-MAR-2000 (first entry)

DE Neisseria meningitidis ORF 212 partial DNA sequence SEQ ID NO:735.

KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
KW antibacterial; gene therapy; ds.

OS Neisseria meningitidis.

PN MO9957280-A2.

XX 11-NOV-1999.

PD 30-APR-1999; 99WO-US09346.

XX 01-MAY-1998; 98US-0083758.

PR 31-JUL-1998; 98US-0094869.

PR 02-SEP-1998; 98US-0098994.

PR 02-SEP-1998; 98US-0099062.

PR 09-OCT-1998; 98US-0103749.

PR 09-OCT-1998; 98US-0103794.

PR 09-OCT-1998; 98US-0103796.

PR 25-FEB-1999; 99US-0121528.

XX
XX
PA (CHIR) CHIRON CORP.
PA (GENO-) INST GENOMIC RES.

PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
PI Petersen J, Pizsa M, Rappelli R, Ratti G, Scalato E, Scarselli M;
PI Tettelin H, Venter JC;
XX
XX
DR WPI: 2000-062150/05.
DR P-PSDB; AAY74631.

PT Novel Neisserial polypeptides predicted to be useful antigens for
PT vaccines and diagnostics -
XX
XX
PS Claim 7; Page 488-489; 1453pp; English.

CC AA253015 to AA254536, AA254577 to AA254615, and AAY74253 to AAY75941
CC represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
CC and polypeptides. AA254537 to AA254576 and AA254616 to AA254773 represent
CC PCR primers used in the exemplification of the present invention. The
CC polypeptides, the polynucleotides, antibodies and compositions of
CC the invention can be used as vaccines, as diagnostic reagents, and as
CC immunogenic compositions. The polypeptides can be used in the
CC manufacture of medicaments for treating or preventing infection due to
CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the
CC presence of Neisseria bacteria, or to raise antibodies. They may also
CC be used to screen for agonists or antagonists, which may themselves
CC have use as antibacterial agents. The polynucleotides of the invention
CC may also be used in gene therapy protocols.

XX
XX
SQ Sequence 1620 BP; 412 A; 625 C; 318 G; 260 T; 5 other;

Query Match 29.1%; Score 29.4; DB 21; Length 1620;
Best Local Similarity 53.7%; Pred. No. 20;
Matches 51; Conservative 4; Mismatches 40; Indels 0; Gaps 0;

OY 6 CCGGATGATACCGCGGTGCGGTGATGCGGACCGAGGACGACGAGCGCGCT 65

Db 1588 CGCGATGATGTCGCGGTTGGGCTGTAGCGGTGCGAGCGGTTGTAAGAAGCGATGG 1529
OY 66 GCAATTCACGTTGCGAGCGGCGGCGGCGGCGAGC 100
Db 1528 GCARCGCTASGATTTGGGCTGCAMTGCAGCGCGG 1494

RESULT 12

AAV58938
ID AAV58938 standard; DNA; 10095 BP.

XX AAV58938;

XX 15-FEB-1999 (first entry)

XX Mycobacterium tuberculosis embCAB operon.

XX Drug resistance; antibiotic resistance; antimycobacterial;

KW ethambutol; embCAB operon; infection; vaccine; therapy;

KW tuberculosis; ds.

OS Mycobacterium tuberculosis.

XX WO9841533-A1.

XX 24-SEP-1998.

XX 16-MAR-1998; 98WO-US05128.

XX 20-MAR-1997; 97US-0822586.

XX (YESH) UNIV YESHIVA EINSTEIN COLLEGE.

XX Jacobs WR, Musser J, Telenti A;

XX WPI; 1998-521160/44.

XX P-PSDB; AAM73052-54.

XX Wild type and mutated sequences of Mycobacterium embCAB operon -
PT useful to e.g. identify ethambutol-resistant mycobacterial strains
PS and produce antisense sequences to treat mycobacterial infections
XX Disclosure; Fig 4A-I; 62pp; English.

XX This is the DNA sequence of the Mycobacterium tuberculosis embCAB
CC operon, which determines resistance to the antimycobacterial drug
CC ethambutol (EMB). It includes the embA, embC and embB genes
CC that encode proteins (see AAM73052-54) which are the target of
CC action of M. tuberculosis for EMB. Wild-type and mutated embCAB
CC nucleic acid sequences are useful e.g. as probes used in the
CC diagnosis of drug-resistant mycobacteria or to determine the
CC susceptibility of mycobacteria to EMB. The nucleic acids are also
CC useful in the treatment of mycobacterial infections; anti-DNA or
CC anti-RNA sequences can be administered to inhibit embCAB operon
CC mRNA activity (claimed). The invention additionally provides for
CC the use of embCAB operon nucleic acid sequences as vaccines, or
CC to improve existing vaccines.

XX Sequence 10095 BP; 1469 A; 3475 C; 3370 G; 1780 T; 1 other;

XX Query Match 28.7%; Score 29; DB 19; Length 10095;

XX Best Local Similarity 55.4%; Pred. No. 30;

XX Matches 56; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

OY 1 GCCTTCGGGATGATACCGCGGTGCGAGTATGCGACCGAGGACGACCGC 60

OY 61 GCGCTGCAATTCACGTTGCGAGCGGCGGCGGCGGCGGCGGCGG 101

OY 925 CACCGGCGGTTCTGCGCGCGCTGTGTGTCGACCGCGG 965

RESULT 13
ID AAF88338
XX AAF88338 standard; DNA; 14775 BP.

XX AAF88338;

XX 28-AUG-2001 (first entry)

XX S. spinosa DNA fragment encoding ORF21, SEQ ID 47.

XX Forosamine; trimethylrhamsone; polyketide synthase; biosynthesis;

KW spinosyn; polyketide aglycone; transgenic plant; insect resistance;

KW macrolide; insecticidal; polyketide synthase; ds.

XX Saccharopolyspora spinosa.

XX DE19957268-A1.

XX 08-MAR-2001.

XX 29-NOV-1999; 99DE-1057268.

XX 27-AUG-1999; 99DE-1040596.

XX (FARB) BAYER AG.

XX Eberz G, Moehrl V, Froede R, Velten R, Salas JA;

XX WPI; 2001-267102/28.

XX P-PSDB; AAB70968.

XX New nucleic acid encoding enzymes for spinosyn biosynthesis, useful for
XX recombinant production of insecticidal spinosyns and their derivatives

XX Claim 7a; Page 239-264; 354pp; German.

XX This invention describes a novel method nucleic acid (I) and its encoded
CC polypeptide (II) containing at least one region that encodes an enzymatic
CC activity involved in biosynthesis of spinosyns. (I) are used (i) to
CC identify, inactivate or modulate genes involved in the biosynthesis of
CC (II); (ii) to generate a library of polyketide synthases; (iii) for
CC adding forosamine or trimethylrhamsone to a spinosyn or polyketide
CC aglycone; and (iv) for recombinant production of the corresponding
CC enzymes, which are used for production of (II), their precursors or
CC derivatives, including production of transgenic plants that express (II)
CC and thus have increased resistance to insects. (I) are also useful as
CC markers for sequencing of the Saccharopolyspora spinosa genome. (II) are
CC also used to raise specific antibodies, useful for identifying
CC expression clones in a gene bank. Cells transformed with (I) may produce
CC (II) at significantly increased levels or produce new derivatives of
CC (II). This sequence, ORF 21, encodes an S. spinosa polyketide synthase.

XX Sequence 14775 BP; 1969 A; 4181 C; 5731 G; 2894 T; 0 other;

XX Query Match 28.5%; Score 28.8; DB 22; Length 14775;

XX Best Local Similarity 58.0%; Pred. No. 36;

XX Matches 51; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

OY 6 CGGGATGATACCGCGGCGGCGGTGCGAGTATGCGCGGAGGAGGAGCGCGGCT 65

OY 4095 CGTGACGCGCACCTGTGAGTGTGCGAGCGCTTGCGGCGCGGTGCGAGGAGAAC 4134

OY 66 GCAATTCACGTTGCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGG 93

OY 4155 GCAGCTGCACTTTCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4182

RESULT 14
ID AAF88313
XX AAF88313 standard; DNA; 50000 BP.

```

XX AF88313;
AC 28-AUG-2001 (first entry)
XX
DE S. spinosa DNA fragment SEQ ID 2.
XX
XX Forosamine; trimethylirhamnose; polyketide synthase; biosynthesis;
XX spinosyn; polyketide aglycone; transgenic plant; insect resistance;
XX macrolide; insecticidal; ds.
XX
XX Saccharopolyspora spinosa.
XX
XX DE19957268-A1.
XX
XX 08-MAR-2001.
XX
XX 29-NOV-1999; 99DE-1057268.
XX
XX 27-AUG-1999; 99DE-1040596.
XX
XX (FARB ) BAYER AG.
XX
XX Eberz G, Moehle V, Froede R, Velten R, Salas JA;
XX
XX WPI; 2001-267102/28.
XX
XX New nucleic acid encoding enzymes for spinosyn biosynthesis, useful for
XX recombinant production of insecticidal spinosyns and their derivatives
XX
XX
XX Claim 7; Page 31-49; 354pp; German.
XX
XX This invention describes a novel method nucleic acid (I) and its encoded
XX polypeptide (II) containing at least one region that encodes an enzymatic
XX activity involved in biosynthesis of spinosyns. (I) are used (i) to
XX identify, inactive or modulate genes involved in the biosynthesis of
XX (II); (ii) to generate a library of polyketide synthases; (iii) for
XX adding forosamine or trimethylirhamnose to a spinosyn or polyketide
XX aglycone; and (iv) for recombinant production of the corresponding
XX enzymes, which are used for production of (II), their precursors or
XX derivatives, including production of transgenic plants that express (II)
XX and thus have increased resistance to insects. (I) are also useful as
XX markers for sequencing of the Saccharopolyspora spinosa genome. (II) are
XX macroclides with insecticidal, but not antibacterial, activity, and can
XX also be used to raise specific antibodies, useful for identifying
XX expression clones in a gene bank. Cells transformed with (I) may produce
XX (II) at significantly increased levels or produce new derivatives of
XX (II). This sequence represents a genomic DNA fragment of the S. spinosa
XX genome which contains the coding regions for proteins involved in
XX forosamine, trimethylirhamnose and polyketide synthase biosynthesis.
XX
XX Sequence 50000 BP; 6867 A; 14165 C; 19274 G; 9694 T; 0 other:
XX
XX
XX Query Match 28.5%; Score 28.8; DB 22; Length 50000;
XX Best Local Similarity 58.0%; Pred. No. 40;
XX Matches 51; Conservative 0; Mismatches 37; Indels 0; Gaps 0;
XX
XX 6 CCGGATGATACCGCGGTGGCGGATGCGACCGAGAGAGACGAGCCGCT 65
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX Db 23701 CCGGACGGACACACGAGTCGTGGCAGCGGCTGCCGCCGTGGACAGAGAAC 23760
XX
XX 66 GCAATTCACAGTTCGAGCGCGGTGGCG 93
XX ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX Db 23761 GCAGCTGGCACCTTCGGCGCGGCGTGGCG 23788
XX
XX
XX RESULT 15
XX AAF88316
XX ID AAF88316 standard; DNA: 50000 BP.
XX
XX AC AAF88316;
XX

```

```

DT 28-AUG-2001 (first entry)
XX
XX S. spinosa DNA fragment SEQ ID 5.
XX
XX Forosamine; trimethylirhamnose; polyketide synthase; biosynthesis;
XX spinosyn; polyketide aglycone; transgenic plant; insect resistance;
XX macrolide; insecticidal; ds.
XX
XX Saccharopolyspora spinosa.
XX
XX DE19957268-A1.
XX
XX 08-MAR-2001.
XX
XX 29-NOV-1999; 99DE-1057268.
XX
XX 27-AUG-1999; 99DE-1040596.
XX
XX (FARB ) BAYER AG.
XX
XX Eberz G, Moehle V, Froede R, Velten R, Salas JA;
XX
XX WPI; 2001-267102/28.
XX
XX New nucleic acid encoding enzymes for spinosyn biosynthesis, useful for
XX recombinant production of insecticidal spinosyns and their derivatives
XX
XX
XX Claim 7; Page 74-91; 354pp; German.
XX
XX This invention describes a novel method nucleic acid (I) and its encoded
XX polypeptide (II) containing at least one region that encodes an enzymatic
XX activity involved in biosynthesis of spinosyns. (I) are used (i) to
XX identify, inactive or modulate genes involved in the biosynthesis of
XX (II); (ii) to generate a library of polyketide synthases; (iii) for
XX adding forosamine or trimethylirhamnose to a spinosyn or polyketide
XX aglycone; and (iv) for recombinant production of the corresponding
XX enzymes, which are used for production of (II), their precursors or
XX derivatives, including production of transgenic plants that express (II)
XX and thus have increased resistance to insects. (I) are also useful as
XX markers for sequencing of the Saccharopolyspora spinosa genome. (II) are
XX macroclides with insecticidal, but not antibacterial, activity, and can
XX also be used to raise specific antibodies, useful for identifying
XX expression clones in a gene bank. Cells transformed with (I) may produce
XX (II) at significantly increased levels or produce new derivatives of
XX (II). This sequence represents a genomic DNA fragment of the S. spinosa
XX genome which contains the coding regions for proteins involved in
XX forosamine and trimethylirhamnose biosynthesis.
XX
XX Sequence 50000 BP; 6944 A; 14312 C; 19170 G; 9574 T; 0 other:
XX
XX
XX Query Match 28.5%; Score 28.8; DB 22; Length 50000;
XX Best Local Similarity 58.0%; Pred. No. 40;
XX Matches 51; Conservative 0; Mismatches 37; Indels 0; Gaps 0;
XX
XX 6 CCGGATGATACCGCGGTGGCGGATGCGACCGAGAGAGACGAGCCGCT 65
XX ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX Db 28077 CCGGACGGACACACGAGTCGTGGCAGCGGCTGCCGCCGTGGACAGAGAAC 28136
XX
XX 66 GCAATTCACAGTTCGAGCGCGGTGGCG 93
XX ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX Db 28137 GCAGCTGGCACCTTCGGCGCGGCGTGGCG 28164
XX
XX
XX Search completed: November 6, 2002, 12:46:48
XX Job time : 1760 secs

```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 6, 2002, 11:51:49 ; Search time 443.625 Seconds

(without alignments)
6625.829 Million cell updates/sec

Title: US-09-724-876-2_COPY_34350_34450

Perfect score: 101

Sequence: 1 gcttcgcgatgataccgc.....agccgggtgcgggccaacg 101

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 segs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:*

- 1: gb_ba:*
- 2: gb_htg:*
- 3: gb_in:*
- 4: gb_om:*
- 5: gb_ov:*
- 6: gb_pat:*
- 7: gb_ph:*
- 8: gb_pl:*
- 9: gb_pr:*
- 10: gb_ro:*
- 11: gb_sts:*
- 12: gb_sy:*
- 13: gb_un:*
- 14: gb_vi:*
- 15: em_ba:*
- 16: em_fun:*
- 17: em_hum:*
- 18: em_in:*
- 19: em_mu:*
- 20: em_om:*
- 21: em_or:*
- 22: em_ov:*
- 23: em_pat:*
- 24: em_ph:*
- 25: em_pl:*
- 26: em_ro:*
- 27: em_sts:*
- 28: em_un:*
- 29: em_vi:*
- 30: em_htg_hum:*
- 31: em_htg_inv:*
- 32: em_htg_other:*
- 33: em_htg_mus:*
- 34: em_htg_pln:*
- 35: em_htg_rtd:*
- 36: em_htg_mam:*
- 37: em_htg_vrt:*
- 38: em_sy:*
- 39: em_htgo_hum:*
- 40: em_htgo_mus:*
- 41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	101	100.0	58733	1 AF217189	AF217189 Sorangium
2	101	100.0	71989	6 AF217264	AF217264 Sequence
3	99.4	98.4	68750	1 AF210843	AF210843 Sorangium
4	99.4	98.4	68750	6 AR193029	AR193029 Sequence
5	99.4	98.4	68750	6 AR199551	AR199551 Sequence
6	99.4	98.4	68750	6 AR199559	AR199559 Sequence
7	99.4	98.4	68750	6 AR199567	AR199567 Sequence
8	99.4	98.4	68750	6 AR201097	AR201097 Sequence
9	99.4	98.4	68750	6 AR208671	AR208671 Sequence
10	65.2	64.6	16124	1 AX024384	AX024384 Sequence
11	65.2	64.6	16124	6 AX024277	AX024277 Sequence
12	34.4	34.1	609	4 AY112896	AY112896 Equus cab
13	31.8	31.5	10507	1 AE012031	AE012031 Xanthomon
14	31.6	31.3	17205	1 AE007185	AE007185 Mycobacte
15	31.6	31.3	23740	1 MTW026	MTW026 Mycobacte
16	31.6	31.3	96774	2 AP004163	AP004163 Oryza sat
17	31.6	31.3	148766	2 AC118895	AC118895 Rattus no
18	31.4	31.1	37445	1 SC8A6	SC8A6 Rattus no
19	31.2	30.9	11910	1 AE012015	AE012015 Xanthomon
20	31.2	30.9	310050	1 RME603642	RME603642 Rhizobium
21	31	30.7	409	14 AF065707	AF065707 Stealth v
22	31	30.7	845	14 AF065691	AF065691 Stealth v
23	31	30.7	215050	2 AL646057	AL646057 Ralstonia
24	30.8	30.5	235571	2 AC128364	AC128364 Rattus no
25	30.6	30.3	2947	1 SL118818	SL118818 Streptomyce
26	30.6	30.3	3012	10 AF293383	AF293383 Rattus no
27	30.6	30.3	4851	1 SC018817	SC018817 Streptomyce
28	30.6	30.3	12647	1 AE012088	AE012088 Xanthomon
29	30.6	30.3	29080	1 SCBAC28G1	SCBAC28G1 Streptomy
30	30.6	30.3	40337	1 SC5A7	SC5A7 Streptomy
31	30.6	30.3	140714	2 RN374E16	RN374E16 Rattus no
32	30.6	30.3	190050	1 AL646059	AL646059 Ralstonia
33	30.4	30.1	1078	3 AY051416	AY051416 Drosophila
34	30.4	30.1	106186	8 AF448416	AF448416 Zea mays
35	30.4	30.1	123751	2 AC019937	AC019937 Drosophila
36	30.4	30.1	153749	8 AP003019	AP003019 Oryza sat
37	30.4	30.1	156508	3 AC007581	AC007581 Drosophila
38	30.4	30.1	171151	3 AC007574	AC007574 Drosophila
39	30.4	30.1	226001	8 AF391808	AF391808 Zea mays
40	30.4	30.1	310958	3 AE003464	AE003464 Drosophila
41	30.2	29.9	11169	1 AE004741	AE004741 Pseudomon
42	30.2	29.9	84487	2 OSJN00006	OSJN00006 Oryza sat
43	30.2	29.9	115548	2 OSJN00010	OSJN00010 Oryza sat
44	30	29.7	2137	9 AB060236	AB060236 Macaca fa
45	30	29.7	2212	1 BS068411	BS068411 Burkholderia

ALIGNMENTS

RESULT 1
AF217189
LOCUS
DEFINITION
AF217189
58733 bp DNA linear BCT 09-JUN-2000
Sorangium cellulosum putative transposase gene, partial cds;
putative transposase gene, complete cds; epothilone biosynthesis
gene cluster, complete sequence; putative membrane protein gene,
complete cds.
ACCESSION
AF217189
VERSION
AF217189.1 GI:7453554
KEYWORDS
ORGANISM
REFERENCE
1 (bases 1 to 58733)
Sorangium cellulosum.
Polyangium cellulosum.
Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
Myxococcales; Sorangium; Polyangium; Polyangiaceae; Polyangium.

QSTRMWTETLIDLRPLMGLDHRVQAVVFPGAALYEMALISSGAELAD3PLQITDV
LAELAFAGDAVAVOYVTTTEPQSGRLOFQIASRPGGHSFVPHAGALLVRETE
VPAGTILSAVARLQASMPAAATAYELTEMLOIYASRPGALFELRMEGELGVRRE
DAAGSAARLRLHPLDLACROYVCSLFRAGGEATPWPEVSGIRLLOPSEGLMCHA
RVNHGQTPDROGADFVVDSSGAVVAEVSGLVAQRI.PGVGVRREEDDWLEKMEP
AAVTGAVNAGRMRLILGGGGGLGAALRSMLEAGAAVHAASNTSSAAGVALLAKAF
DQAPTAIVHLISDGGGELDPGLAGGALDAPRSADVSPALDPALVRCGDSYLVTV
OALAGMGRDAPRLMLTLRGAAVAGDVSVTOAPLGLIGRTVIMEHADLCARVLD
PTRPDGLGALLAELDLSTAGVLAERAGHLVVGSGAASVQDRAVALERGAR
IRADSTLVYTGGLGGLSTAGVLAERAGHLVVGSGAASVQDRAVALERGAR
VTAKAVADADQLELRLREVTTSGMPLRGVHAAGLIDDLMOQPPARRKWAK
VOGALHHLATREAPLSEFVLVYASGVGLSGGGGNTAAATFEDALAHHRAGGLPA
LSVDMGLFAEYGAQAQEDRGARLVSRGMSRLPTDEGLSAAATFLESLVRGGDVLPV
PRLMVELYPAASRMLSRVTGARSRGAGGDDGLRLRAAEPSARSGLEPLLR
AOSIOVRLPBGKTEVDAPLISGMNSLMGLENRRIEAMGITVPAFLMTYPTVAA
LSGHLARACEAARVESPHITADSARETEMSQDULIOLIAKTRALT"

gene
16134..37907
/gene="epod"
16134..37907
/gene="epod"
/note="type 1 polyketide synthase; epothilone PKs modules
3, 4, 5, and 6"
/codon_start=1
/transl_table=1
/product="epod"
/protein_id="AAF62883.1"
/db_xref="GI:7453560"
/translation="MTTRGPPAQONPLKOAIIITOREBERLAGIAQALEPTEPIAY
GICRFPFGADAPAEFWELDLAERDAVOPDLDRNALVGAVAEAVPHMAGLTPTDC
FDAFPGIISPEARSLDPOHRLILEVAMEGLDEDIPRSIDGSRTEGVGATFADYA
RTVARLREERDAYSATGNMILSIAGRSYTLGOGPCLVYDITACSSLVLIHACRS
LRAGESDLALAGVSTLLSPDMEAAARTOALSPDRCRTFDANANGVREGGGLVY
LKRLSDAORODRITWALIRGSAIHDGRSGTCLPNTLAQCTIVLRALRSVHAGAV
DYVTRHGTSGISLDPTEYERALRATVGPARSQTCVYCAVATNIGHLEBAAGVGLIK
AASLTERIRIPRINERTLNRIIRLESALALATEPVWPPTDPRRAGVSTFMSGT
NAHVLEAPAVELMPAPAPERSABELVLSGSEBALDAPARTDREHDMHELIGDY
AFSLATRRASATHLAAVATRSREGILALSAVAGOTPAQAARCIASSRKLALLET
GOGAOTGGMGGLCAAMPAPREAREDCRTLEDRLDRLRPMVAEAGSARSLLDOR
AFQGPALFAVEYALTAIAMSWSGVPELLVSGEIVLAACVAFSELDGCRVLAAR
RLMGSLAGGAMSLGAPAEVAVAAPAHAMVSIATAVNGEYQVYACVEDAVOALIA
GFARGVRTKRLHVSHPMLPEERGERVAASVYTRRPSVSLVNSLKGVTDE

Query Match 100.0%; Score 101; DB 1; Length 58733;
Best Local Similarity 100.0%; Pred. No. 2,8e-15;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCTTCGGGATGATACCGCGGTGGGTGCAGTGCAGCGAGGACGACGAC 60
|||||
Db 34350 GCCTTCGGGATGATACCGCGGTGGGTGCAGTGCAGCGAGGACGACGAC 609
|||||

QY 61 CGCCTGCAATTCACGTCGCGCGGTGCCGGCGGCGCACGG 101
|||||
Db 34410 CGCCTGCAATTCACGTCGCGCGGTGCCGGCGGCGCACGG 34450
|||||

RESULT 2
AR172664 71989 bp DNA linear PAT 17-DEC-2001
DEFINITION Sequence 2 from patent US 6303342.
ACCESSION AR172664
VERSION AR172664.1 GI:17912155
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 71989)
AUTHORS Julien,B., Katz,L., Khosla,C. and Tang,L.
TITLE Recombinant methods and materials for producing epothilones C and D
JOURNML Patent: US 6303342-A 2 16-OCT-2001;
FEATURES
source Location/Qualifiers
1..71989
/organism="unknown"
BASE COUNT 10108 a 23531 c 26617 g 11731 t 2 others

ORIGIN

Query Match 100.0%; Score 101; DB 6; Length 71989;
Best Local Similarity 100.0%; Pred. No. 2.7e-15;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCTTCGGGATGATACCGCGGTGGGTGCCGGTGCAGTCATGCGACGAGGACGACGAC 60
|||||
Db 34350 GCCTTCGGGATGATACCGCGGTGGGTGCCGGTGCAGTCATGCGACGAGGACGACGAC 34409
|||||

QY 61 CGCCTGCAATTCACGTCGCGCGGTGCCGGCGGCGCACGG 101
|||||
Db 34410 CGCCTGCAATTCACGTCGCGCGGTGCCGGCGGCGCACGG 34450
|||||

RESULT 3
AF210843 68750 bp DNA linear BCT 21-JAN-2000
LOCUS Sorangium cellulosum strain So ce90 epothilone biosynthesis gene
DEFINITION cluster, complete sequence.
ACCESSION AF210843
VERSION AF210843.1 GI:6724237
KEYWORDS
SOURCE Polyangium cellulosum.
ORGANISM Polyangium cellulosum.
Bacteria; Proteobacteria; Polyangiaceae; Polyangium.
Myxococcales; Sorangineae; Polyangiaceae; Polyangium.
Molnar,I., Schupp,T., Ono,M., Zikkle,R., Milnnow,M.,
Nowak-Thompson,B., Engel,N., Toupet,C., Stratanan,A., Cyr,D.D.,
Gorlach,J., Mayo,J.M., Hu,A., Goff,S., Schmid,J. and Ligon,J.M.
The biosynthetic gene cluster for the microtubule-stabilizing
agents epothilones A and B from Sorangium cellulosum So ce90
Chem. Biol. 7 (2), 97-109 (2000)

JOURNAL
MEDLINE 20130945
PUBMED 10662695
REFERENCE 1 (bases 1 to 68750)
AUTHORS Molnar,I.
TITLE Direct Submission
JOURNML Submitted (03-DEC-1999) Natural Product Genetics, Novartis
Agribusiness Research Institute, Inc., 3054 Cornwallis Rd, P.O. Box
12257, Research Triangle Park, NC 27709, USA
FEATURES
source Location/Qualifiers
1..68750
/organism="Polyangium cellulosum"
/strain="So ce90"
/db_xref="taxon:56"
<1..1826
/note="ORF1"
/codon_start=3
/transl_table=1
/product="unknown"
/protein_id="AAF26904.1"
/db_xref="GI:6724238"
/translation="ASIDALFARATSARVLDDGGRATERHVLAENGIEDLRALREH
LRIGSGSPHCMLGIDLTVELLAHDDPLASISPHNARSLEPHWTDNAMIYDPAIY
RWLAARGAPRLREYEREREAKTQDERRMILAAAPCPAPDLPREDPANGPLPG
MSPVAEERLRASVATPELACALALAWITGAGPWSGYATYMLPENLILGGLTP
AIAAASAPGTSEALRGAARLPASMEVVSRSKQGNIPALMERLRITYAMNADN
LSREFERAIAIAEVRLLAOPAPPAAGGLAVAGVSSGSGISGLVTGDALYSGDND
IVMFOGRISPVNVLAGTDPPEFLAPLSOULFAVHANAAGTISKVLEGGSLIYARN
QARMSIYHARGFMVANOAVPDPBERGAPRVYORSTIMEERHTPCRIHPAGSARS
LACDEHILWELSGRLERLRRHHRHARGARSRAYIGENHIAATWIPSLITANTHYL
WADPDRRLILGVDRKTGVEPTVLAEIRHPAHVSEDDIETALTGQDPSRDMVHEHR
SGASTVADYOROLMDRDWVLAETRLNRRGLEFTTNDRLTLARS"
complement(1900..3171)
/note="ORF2: similar to hypothetical proteins from
Mycobacterium tuberculosis and Streptomyces coelicolor and
to DD-peptidases"
/codon_start=1
/transl_table=1
/product="unknown"
/protein_id="AAF26905.1"

Best Local Similarity 99.0%; Pred. No. 6.7e-15;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 GCCTTCGGGATGATACCGCGGTGCGGTCATGCGCAGGAGGAGGACGAGC 60
|||||
Db 39962 GCCTTCGGGATGATACCGCGGTGCGGTCATGCGCAGGAGGAGGACGAGC 40021
|||||

Oy 61 CGCCTGCAATTCACGTTGCGAGCGGGTGCCGGGCCACGG 101
|||||
Db 40022 CGCCTGCAATTCACGTTGCGAGCGGGTGCCGGGCCACGG 40062
|||||

RESULT 4
AR193029 68750 bp DNA linear PAT 20-APR-2002
LOCUS AR193029
DEFINITION Sequence 1 from patent US 6346404.
ACCESSION AR193029
VERSION AR193029.1 GI:20238994
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
BASE COUNT 9596 a 22456 c 25539 g 11159 t
ORIGIN

Query Match 98.4%; Score 99.4; DB 6; Length 68750;
Best Local Similarity 99.0%; Pred. No. 6.7e-15;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 GCCTTCGGGATGATACCGCGGTGCGGTCATGCGCAGGAGGAGGACGAGC 60
|||||
Db 39962 GCCTTCGGGATGATACCGCGGTGCGGTCATGCGCAGGAGGAGGACGAGC 40021
|||||

Oy 61 CGCCTGCAATTCACGTTGCGAGCGGGTGCCGGGCCACGG 101
|||||
Db 40022 CGCCTGCAATTCACGTTGCGAGCGGGTGCCGGGCCACGG 40062
|||||

RESULT 5
AR199551 68750 bp DNA linear PAT 20-APR-2002
LOCUS AR199551
DEFINITION Sequence 1 from patent US 6355457.
ACCESSION AR199551
VERSION AR199551.1 GI:20249625
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
BASE COUNT 9596 a 22456 c 25539 g 11159 t
ORIGIN

Query Match 98.4%; Score 99.4; DB 6; Length 68750;
Best Local Similarity 99.0%; Pred. No. 6.7e-15;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 GCCTTCGGGATGATACCGCGGTGCGGTCATGCGCAGGAGGAGGACGAGC 60
|||||
Db 40022 CGCCTGCAATTCACGTTGCGAGCGGGTGCCGGGCCACGG 40062
|||||

Db 39962 GCCTTCGGGATGATACCGCGGTGCGGTCATGCGCAGGAGGAGGACGAGC 40021
Oy 61 CGCCTGCAATTCACGTTGCGAGCGGGTGCCGGGCCACGG 101
|||||
Db 40022 CGCCTGCAATTCACGTTGCGAGCGGGTGCCGGGCCACGG 40062
|||||

RESULT 6
AR199559 68750 bp DNA linear PAT 20-APR-2002
LOCUS AR199559
DEFINITION Sequence 1 from patent US 6355458.
ACCESSION AR199559
VERSION AR199559.1 GI:20249633
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
BASE COUNT 9596 a 22456 c 25539 g 11159 t
ORIGIN

Query Match 98.4%; Score 99.4; DB 6; Length 68750;
Best Local Similarity 99.0%; Pred. No. 6.7e-15;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 GCCTTCGGGATGATACCGCGGTGCGGTCATGCGCAGGAGGAGGACGAGC 60
|||||
Db 39962 GCCTTCGGGATGATACCGCGGTGCGGTCATGCGCAGGAGGAGGACGAGC 40021
|||||

Oy 61 CGCCTGCAATTCACGTTGCGAGCGGGTGCCGGGCCACGG 101
|||||
Db 40022 CGCCTGCAATTCACGTTGCGAGCGGGTGCCGGGCCACGG 40062
|||||

RESULT 7
AR199567 68750 bp DNA linear PAT 20-APR-2002
LOCUS AR199567
DEFINITION Sequence 1 from patent US 6355459.
ACCESSION AR199567
VERSION AR199567.1 GI:20249641
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
BASE COUNT 9596 a 22456 c 25539 g 11159 t
ORIGIN

Query Match 98.4%; Score 99.4; DB 6; Length 68750;
Best Local Similarity 99.0%; Pred. No. 6.7e-15;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 GCCTTCGGGATGATACCGCGGTGCGGTCATGCGCAGGAGGAGGACGAGC 60
|||||
Db 39962 GCCTTCGGGATGATACCGCGGTGCGGTCATGCGCAGGAGGAGGACGAGC 40021
|||||

Oy 61 CGCCTGCAATTCACGTTGCGAGCGGGTGCCGGGCCACGG 101
|||||
Db 40022 CGCCTGCAATTCACGTTGCGAGCGGGTGCCGGGCCACGG 40062
|||||

RESULT 8
LOCUS AR201097 68750 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 1 from patent US 6358719.
ACCESSION AR201097
VERSION AR201097.1 GI:20251985
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schnpp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and Goriach,J.
JOURNAL Genes for the biosynthesis of epoethliones
PATENT: US 6358719-A 1 19-MAR-2002;
FEATURES
source 1. 68750 /organism="unknown"
BASE COUNT 9596 a 22456 c 25539 g 11159 t
ORIGIN
Query Match 98.4%; Score 99.4; DB 6; Length 68750;
Best Local Similarity 99.0%; Pred. No. 6.7e-15;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 GCCTGCGGATGATACGCCGGTGGCGGTGCAGTGCAGCGACGAGACGACGAGCC 60
DB 39962 GCCTGCGGATGATACGCCGGTGGCGGTGCAGTGCAGCGACGAGACGAGCC 60
QY 61 CGCCTGCAATTCCACGTTGGAGCCGGGTGCCGGCCACGG 101
DB 40022 CGCCTGCAATTCCACGTTGGAGCCGGGTGCCGGCCACGG 40062
RESULT 10
LOCUS AX024384 16124 bp DNA linear BCT 15-SEP-2000

DEFINITION Sequence 82 from Patent DE19846493.
ACCESSION AX024384
VERSION AX024384.1 GI:10184588
KEYWORDS
SOURCE Polyangium cellulosum.
ORGANISM Polyangium cellulosum
REFERENCE 1 (bases 1 to 16124)
AUTHORS Myxococcales; Sorangineae; delta subdivision; Myxobacteria;
JOURNAL Beyer,S. and Mueller,R.J.
PATENT: DE 19846493-A 13-APR-2000;
FEATURES
source 1. 16124 /organism="Polyangium cellulosum"
BASE COUNT 2336 a 5400 c 5695 g 2693 t
ORIGIN
Query Match 64.6%; Score 65.2; DB 1; Length 16124;
Best Local Similarity 80.9%; Pred. No. 1.8e-06;
Matches 76; Conservative 0; Mismatches 18; Indels 0; Gaps 0;
QY 2 CCTTCGCGATGATACGCCGGTGGCGGTGCAGTGCAGCGACGAGACGACGAGCC 61
DB 2638 CCTTCGCGCGATACGCCGGTGGCGGTGCAGTGCAGCGACGAGACGAGCC 2697
QY 62 GCCTGCAATTCCACGTTGGAGCCGGGTGCCGGCCACGG 95
DB 2698 GCCTGCGGTTCCAGTACGAGTGGGAGCCGGG 2731

RESULT 11
LOCUS AX024277 16124 bp DNA linear PAT 15-SEP-2000
DEFINITION Sequence 82 from Patent DE19846493.
ACCESSION AX024277
VERSION AX024277.1 GI:10184551
KEYWORDS
SOURCE Polyangium cellulosum.
ORGANISM Polyangium cellulosum
REFERENCE 1 (bases 1 to 16124)
AUTHORS Myxococcales; Sorangineae; Polyangineae; Myxobacteria;
JOURNAL Beyer,S. and Mueller,R.J.
PATENT: DE 19846493-A 13-APR-2000;
FEATURES
source 1. 16124 /organism="Polyangium cellulosum"
BASE COUNT 2336 a 5400 c 5695 g 2693 t
ORIGIN
Query Match 64.6%; Score 65.2; DB 6; Length 16124;
Best Local Similarity 80.9%; Pred. No. 1.8e-06;
Matches 76; Conservative 0; Mismatches 18; Indels 0; Gaps 0;
QY 2 CCTTCGCGATGATACGCCGGTGGCGGTGCAGTGCAGCGACGAGACGACGAGCC 61
DB 2638 CCTTCGCGCGATACGCCGGTGGCGGTGCAGTGCAGCGACGAGACGAGCC 2697
QY 62 GCCTGCAATTCCACGTTGGAGCCGGGTGCCGGCCACGG 95
DB 2698 GCCTGCGGTTCCAGTACGAGTGGGAGCCGGG 2731

RESULT 12
LOCUS AY112896 609 bp mRNA linear MM 09-JUN-2002
DEFINITION Equus caballus Indian hedgehog (IHH) mRNA, partial cds.
ACCESSION AY112896
VERSION AY112896.1 GI:21360570

FEATURES		Medical Center Dr, Rockville, MD 20850, USA
source	Location/Qualifiers	
gene	1. .17205	/organism="Mycobacterium tuberculosis CDC1551"
	/strain="CDC1551"	/db.xref="taxon:83331"
gene	/note="clinical strain"	complement(130. .1062)
	/gene="MT3910"	complement(130. .1062)
CDS	/gene="MT3910"	/note="similar to PID:804884; identified by sequence similarity; putative"
gene	/codon_start=1	/transl_table=11
	/product="esterase, putative"	/protein_id="AAK48276.1"
CDS	/db.xref="GI:13883794"	/transl_table=11
	/transl_table=11	APENLWVPSPMGRDIPVAFIAGGPHAVYLLDFNAGPVSVMVNTGNAMNTLAGG ISVAPAGGAYSWYTNWEDSGKQMDPFLSALPDMIAANRGLAPGHAIVGAAGGT GAMALAFHDFRFGAGSMGFLPSTTTNGATAGMOQFGVDINGMGAPQLGFW KMDPWHASLLAQNNTRWVWSPTNPGASDPAMIGQAAEAMGNSRMPFYNQYRSVG HNGHDFPAGSDNGMSWAPQLGAMSGDIVAIR"
gene	complement(1209. .2276)	/gene="MT3911"
	complement(1209. .2276)	/note="similar to SP:P31952 GB:X62398 PID:44564 PID:1491793 PID:225974; identified by sequence similarity; putative"
CDS	/codon_start=1	/transl_table=11
	/product="esterase, putative, antigen 85-A"	/protein_id="AAK48277.1"
gene	/db.xref="GI:13883795"	/transl_table=11
	/transl_table=11	VSGIYGVGTATAGATGSRPLVEYILOVSPSMGRDITKVOFOSGNSALYIDTL RADDESGMDINTPAFEMVYDOSGLSYMPVGGOSFTSDYOPACGACGCTYKWEF LHSEPCMLANRHKVKTGSVYVGLSKAASSALTALYHPQOFYAGANGSLDPSA MGPTLIGLAWDGGYKASDMGKEDPAMORNDPLINVKLIANNRWVYCGNGRP SDLGNNLPKFLFEGFVRTSNIKFQDAYNAGGCHNGVDFPDSGTHSMIYGAQINAM KPDLRALGATPNTGAPDGA"
gene	complement(2520. .4445)	/gene="MT3912"
	complement(2520. .4445)	/note="identified by Glimmer2; putative"
CDS	/codon_start=1	/transl_table=11
	/product="hypothetical protein"	/protein_id="AAK48278.1"
gene	/db.xref="GI:13883796"	/transl_table=11
	/transl_table=11	GLIYRTVRNLNLAGNPGVNOGEREANTSTAMTYLLYGVGWGPPRLVVALAM VLSILGVLMLIGRLYAPLSGRRAIMLPAGALYIYVAPADPMTSGESLIVM YGLIMMMYWCSPDLARPDSCMFLGALAFVAGCSYLVPERPLIGLALIMLILAA RTMRRLYIVLAGGFLPVAYOIFRMGTIGLVSSTALAKDAAGKMSQMYVSNFR PYALWPLVIVSPLGILLMTARRSPFLRVLAADVRAAVOSPAAVAFIVGSGV LOALYIROGDEFWHGVLALPLECLAPGVIPILPDGKFSRETRMVLGALSGL WLGJAGSLMAANSPCMGGDATTYVTSIVDEREYLAQATGHAHPLTAADLYDPRM AVLTANNPEGALLPSGNOMDLVMPRPSGTAQGPAPKPOHAAVEFTMSG GMVGLDVRVIDOIGLYNPLAHTERIKARIGDKRLPDMVYADSPWAKVYTGITG YIDQWYTOAEMALQCPATRAVLSVAPITLHRELSNVLSHTEFTKTRIDRVPRTEL VRCGLVDPGPGPPRE"
gene	complement(4492. .5400)	/gene="MT3913"
	complement(4492. .5400)	/note="identified by Glimmer2; putative"
CDS	/codon_start=1	/transl_table=11
	/product="conserved hypothetical protein"	
gene	/protein_id="AAK48279.1"	/transl_table=11
	/db.xref="GI:13883797"	/transl_table=11
CDS	/db.xref="GI:13883797"	RTDYVEVLSKYSMAFVFSLAASAVLYLVNDYRDYDEADREHPTKRFRIAGVPEMLA YVAIVLVGTVSLAGAMMLTPMLALVMVYVLAQMLAYFGKHQVAVICVSSAYTLR AIAGVAFKIPLSKMPFLIMAFGLFVPAKRSFALYELAKRTGAAIKRSLESTYSL RFVWTLATAVLVCGLMAFERDYSQSMFASVMIPTTIALIRYADVDDGLAGEPED IALDRVLOLIALAMVATVGAAYFG"
gene	complement(5407. .5904)	/gene="MT3914"
	complement(5407. .5904)	/note="identified by match to PFAM protein family HM PF01569"
CDS	/codon_start=1	/transl_table=11
	/product="PAP2 superfamily protein"	/protein_id="AAK48280.1"
gene	/db.xref="GI:13883798"	/transl_table=11
	/transl_table=11	RRRMVYAGAGAFYAAHIAVILIKRVRRORDPRAIVANVDTFSQSPFSAHSTSTA AALIMGATGILPVLVLPMAUSRIILGVHYPSDVAVGALGATVGAIVDSGGHQ RAARR"
gene	complement(5933. .7846)	/gene="MT3915"
	complement(5933. .7846)	/note="identified by Glimmer2; putative"
CDS	/codon_start=1	/transl_table=11
	/product="hypothetical protein"	/protein_id="AAK48281.1"
gene	/db.xref="GI:13883799"	/transl_table=11
	/transl_table=11	SLQIASEVSAFYAFNAPASVYRRRTTKSVLRVGVAGRVYRKAAGARTF VEGHDFGTEDQPAVETEVLVLPFEEGGWVPEITDTAVTLHSGMATSAPGTA NIAGVITPTNRPACVNAKRELTPLDVDOYIGAVIYDQGERKVRDHPPEAAATG GSRSLTHDQDSSGGYSRWYKALNTDCCOITLFPDDIRLEPDSILLVNMHRA KAPLVGQMLINDPESHLMGCVYDRSIFPMWTAAPHAEYDHDFAFYPLDNSSRK LHRLIDVYDNGWTCMIPROVAELDQPLPLFKWDADYGLAAAHGYPVTLPEA AIHMAWMSDMDIDAIDMWAYFHLRNRLVVAAMHMGPAQYIGVRSHLKATLKLACL EYSTVAIONKAIIDFLAGPEHIESILSALPRVIRIKSYPDAAVYVLEASLEPPLIK NKAKRPVNPVLYIGRLARGIMHNLTPANPOHHRPEPNVPTODARWFLICTYGAIV TTADCGVVYRORDRAKMPALMQSLRQQLKRFPEMRIRYRDALPTLSSQKWET ALPAAQDEBEHG"
gene	complement(7843. .9042)	/gene="MT3916"
	complement(7843. .9042)	/note="similar to SP:P37747 GB:U03041 GB:U09876 PID:508242 PID:510253; identified by sequence similarity; putative"
CDS	/codon_start=1	/transl_table=11
	/product="UDP-galactopyranose mutase"	/protein_id="AAK48282.1"
gene	/db.xref="GI:13883800"	/transl_table=11
	/transl_table=11	GNAYSEAPQGTLEVHYGAHLFTSKRWMDYRQFTDTDRHRYFAHNGQAYOF PMGGLVISOFFGKXFTPEQARQLIAEAPAEIDVADAGNLEKASLGRVLYEAFVYG YTAQWQTDKEKLEPAANITRLPVYTYTDNRFSQYGLTDTGATLQMMADHRIE VLKMTQEDVRGQLEPGSPAPAVYVTRPLRDFYAGRGMRIDDEVLYPLIGDQ GTAYMANNNDIVPRTIRHRRHPRDEYDTPDKYIMREYSRAEDDEDEYUPLNTEA DRALATYRRAKASETASRYLREGRLGTYQYLDHMAIASALMNTDNYLAPHLRGV PLIDGA"
gene	9306. .10160	/gene="MT3917"
	9306. .10160	/note="similar to PID:845583 GB:AL123456; identified by sequence similarity; putative"
CDS	/codon_start=1	/transl_table=11
	/product="exported repetitive protein"	

gene
CDS
/protein_id="AAK48283.1"
/db_xref="GI:13883801"
/translation="MPNRRRLSTAMSAVALAAVAPCAFLVYESTETTEREHE
FKQAVLIDPGEIMSAISQSGISGICGIPPPVSLGSLGRLPGLTSPGLTSP
LTPSPGLDPAALGANEIPITTTVGLDPPGADGTYPIIGDTLTCTTSSPYTSTGGGL
VNDVQVAVNELGASQAIDILKGVLPMSIQAVQNGAANAASPPVPIPAANAAPPT
DIRTPVA"
10365..11984
/gene="MT3918"
10365..11984
/gene="MT3918"
/note="Identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK48284.1"
/db_xref="GI:13883802"
/translation="MAATVIVAMIANRPPASSHESPPTNTQLAEQPLIGGGVTV
RELTDTPESLVAITGDLGATSAVRARAKPDDMGVQTEYETEPDRDAGDGYEL
GGLNPGPRSTDVFGVTTTQVAVATRPIDATITOPAPRPNLDLSGLRPAATKE
OPEGONISAILSPQAPPGTOMPTPTATVTAAGOPPIISBAEMGDESLRETEPYD
RGVRAVYVHTAGSNDYSPLESAGIKATVYHSKTIQGWICIDAYANLKYGVVEGS
AGGLTPVGEFHTGSPNRTNRYAMIGNDDVAFPPIQITRTYGRLLGRHUGDDVDPK
SNVDLOSAGSYTTFPGGAIARLPALIFTRDVGNTDCCPNAAYAVWDETRDIAHFD
PPEELIKALEGGAITYQRMQALGMSNLSALCAPSPENDADGARYATFAKAMVWSPVT

Query Match 31.3%; Score 31.6; DB 1; Length 17205;
Best Local Similarity 58.5%; Pred. No. 3e+02;
Matches 55; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 1 GCCTTCGGGATGATACGCCGTGGCGGTGCAGTCATGCGACGAGCAGCAGCCAGCC 60
Db 6505 GCCTGCTGGATATATCGACGTGCATCCAGACAGCAGCATCTTCGCC 6446
QY 61 CGCCTGCAATTCACGCTGGAGCCGGGTCCGG 94
Db 6445 GGCCCGGAGCAGCATCTTTCGATCTGGAATCCG 6412

RESULT 15
MTV026/c 23740 bp DNA linear BCT 03-AUG-2001
LOCUS Mycobacterium tuberculosis H37Rv complete genome; segment 157/162.
DEFINITION AL022076 AL123456
ACCESSION AL022076.1 GI:3256026
VERSION
KEYWORDS
SOURCE
ORGANISM
Mycobacterium tuberculosis H37Rv.
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacteriaceae; Mycobacteriaceae;
Mycobacterium tuberculosis complex.

REFERENCE 1 (bases 1 to 23740)
Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C.,
Harris, D., Gordon, S.V., Eigmeier, K., Gas, S., Barry III, C.E.,
Tekle, A.F., Badcock, K., Basham, D., Brown, D., Chillingworth, T.,
Comor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S.,
Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLaren, J.,
Moule, S., Murphy, L., Oliver, S., Osborne, J., Quail, M.A.,
Rajandream, M.A., Rogers, J., Rutter, S., Seeger, K., Skelton, S.,
Squires, S., Squires, J., Sulston, J.E., Taylor, K., Whitehead, S., and
Barrall, B.G.
Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence
Nature 393 (6685), 537-544 (1998)

JOURNAL MEDLINE
PUBMED 98295987
TITLE 9634230
AUTHORS 2 (bases 1 to 23740)
TITLE Parkhill, J.
JOURNAL Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium
tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome
Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetiome

COMMENT
Molecular Bacteriome, Institut Pasteur, 28 rue du Docteur Roux,
75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk
On Jun 26, 1998 this sequence version replaced gi:2961395.
Notes:

Details of M. tuberculosis sequencing on the Sanger Centre are
available on the World Wide Web.
(URL, <http://www.sanger.ac.uk/Projects/M.tuberculosis/>) CDS have
been renumbered from the original cosmid submissions but the old
gene designations are in brackets after the new gene numbers.
Gene prediction was based on a Hidden Markov Model of TB genes
implemented in TParse (Krogh) supplemented with visual inspection
of positional base preference in codons, especially where there is
an increase in the observed/expected third position G + C.
CAUTION: In some cases we may not have predicted the correct
initiation codon. Where possible we choose an initiation codon
(atg, gtg, or ttt) which is preceded by an upstream ribosome
binding site sequence (optimally 5-13bp before the initiation
codon). If this cannot be identified we choose the most upstream
initiation codon.

FEATURES
source
Location/Qualifiers
1..23740
/organism="Mycobacterium tuberculosis H37Rv"
/strain="H37Rv"
/db_xref="taxon:83332"
<1..>23740
/note="fragment designated v026. Does not represent a
physical clone"
60..64
/note="possible RBS for RV3797"
72..1853
/gene="fadE35"
72..1853
/gene="fadE35"
/note="fadE35" (MTV026.02), len: 593. fadE35, Probable
acyl Co-A dehydrogenase similar to E. coli A1dB protein,
gpL20915|EC0A1DB_3 E. coli (A1dB) gene, comp (541 aa);
FASTA scores: opt: 455 z-score: 561.5 E(): 9.2e-24. 31.18
identity in 514 aa overlap. Also similar to MTG176.2 M.
tuberculosis cosmid (542 aa), A1dB, 31.9% identity in 458
aa overlap

misc-feature
RBS
gene
CDS
/codon_start=1
/transl_table=11
/product="fadE35"
/protein_id="CA17861.1"
/db_xref="GI:2950416"
/db_xref="SPTREMBL:053577"
/translation="MPYDLEAVDKLPSTPEKAKQNTENYRGAMGNIWYITPPTQ
ITMATYLRDELATFAEPHUTRIIGELTGCPTVTRMAETRNPPRLERYTRKGDHSRV
LPESFIQSRKRAVIERQAVRDAAAGVPSIALFAADYLLNQADIGACALATGNN
VNSLVTAYAPDPVREFVILKLSNGEMDEAQLTERAGSDGLALETATRGDVL
LNGFWFASNCAGAEVVLAKPEGADSTRGVAFVLTTRRDSRNCVRIARLAKDL
GTRSAVGEIEFVDAEFLLSGEPADSPSGKGLTMMETNNLRIGTSPALNGA
BRAVESICYGORARARAGALIDIKMRKRIAEVYDEALAVFOFGAANRORP
CLPQRIAYPTKLTCKRGITVYASDAITEHGNGTIEWPARLRDQAVNTIEGPD
NLTCLDVRGIEQTRAHETLRLRLDAVSDDDTTLVLRRLIEDLAATYATKTL
ROLAERARLEPLAOEFVDYVAGALLTEQAMBRARRGTRKARVARTVARYLADGPL
RGIDDCDEALORFDELVAGAFYATQ"

gene
CDS
/gene="RV3798"
1980..3314
/gene="RV3798"
1980..3314
/gene="RV3798"
/note="RV3798" (MTV026.03), len: 444. Putative
transposase identical to IS1557 MTCY373_32,
sp|O10621|Y02X_MYCTU_HYPOTHEICAL_50.2_KD PROTEIN CY37
(444 aa), 99.8% identity in 444aa overlap and similar to
many, e.g. gp|U57649|DBU57649_1 Dibenzo(furan)-degrading
bacterium (163 aa), FASTA scores: opt: 767 z-score: 926.8
E(): 0. 67.3% identity in 168 aa overlap, or to BPIS1001.1
B.parapertussis insertion sequenc (406 aa), 24.9% identity
in 402 aa overlap.
/codon_start=1
/transl_table=11
/product="hypothetical protein RV3798"

/protein_id="CAA17862.1"
/db_xref="GI:2950417"
/db_xref="SWISS-PROT:Q10621"
/translation="MRNRRLRLGLVDRKTVIEDIEFEEDDADGAGARVIAVRPRSA
VLRRGRCGRKASWYDRGAGLRKWRSLDWGIVEVLEAEARVNCPTHPGTVAVPAA
RHHGHTYAFDDTVAMALVAVCSKTAVCELMRIARVTAIYARVADTEKRIIDFANL
RRIIDETISYKRHHRYLTVVVDHDSGRVMAAPGHDKATGLFEDALGAEEAAITTY
SADADMLADVYTERCPDAIOGADPFVMAVAPTEALDVERRRANNDARAIARTEPKK
RGRRKNAPRRPGREBARLRKAGYALAKMPEDITEROSAKLMIATIDRLRYAILL
KESLRHVSFKGEGKQALDWISMAORCRTPVEVLEAKIRKRVAIIDALDGLLSQ
GLESTNFKIRLRLRIAFGRSPQALIALMLTLAGHRPTLPGRHNPISQ"
1980..3311
/gene="RV3798"
/note="IS1557-3rd copy"
/complement(3382..4935)
/gene="accD4"
/complement(<3382..4935)
/gene="accD4"
/note="RV3799c, (MTV026.04c), accD4, len: 517. Probable
Propionyl-CoA carboxylase: Beta chain, similar to many,
e.g. SPIP53003|PCCB_SACER PROPIONYL-COA CARBOXYLASE BETA
CH (546aa), FASTA scores: opt: 1549 z-score: 1841.5 E():
0; 50.7% identity in 519 aa overlap, Also similar to
U00012.17 M. leprae cosmid B1308 (549 aa), 49.8% identity
in 514 aa overlap and to MTCY1_20 M. tuberculosis cosmid
(548 aa), 48.9% identity in 515 aa overlap"
/codon_start=1
/transl_table=1
/product="accD4"
/protein_id="CAA17863.1"
/db_xref="GI:2950418"
/db_xref="SPTREMBL:O53578"
/translation="MTYTPVPLHTTAELRLRLAKEPGCEKAARKDKGIPS
ARARITELVDPGSPMETGALCRTPGDNPALIGDGVVGHGILNRPVGLSHDQTVG
GTVGEMFGRVVARLMEWCAMVGCPIVGINDSGARIDAVTSLAMVGLSHDLG
LVQIISITILKCGAGAVYSPIQDLVAVVRGAGVPTGPIVINDVGEVSLDELG
ADHQASGNIHOVYSEAAVQYRDLFLSPSCFPGKPPVNVGLPEIIGHLELD
SIVPDSNMAYDMHEVLRLTFEDGDFDVAAQACQALITGVARNDGTGTVAAQAPPH
MSGALDNEASDARKARTFESDAPDIPVYVAGCPITGVGKRGDITIKKGRRLYAV
VEADVPKVTITIRKSTGATVAGSKULFDLNPMTARIANVIGAGDAGMLKKRP
DPNAPQADAIRKSEVENYNLMALPMTIAERGLDAVIDPHETRLLRKSMILLRDKQ
LMWRVGRKHG"
complement(4932..10133)
/gene="pks13"
complement(4932..10133)
/gene="pks13"
/note="RV3800c, (MTV026.05c), len: 1733. Probable
polyketide synthase, similar to many, e.g.
gplM63676|SERERYAA_1 S. erythraea first ORF of eryA gene,
involved in complex polyketide formation in erythromycin
biosynthesis. FASTA scores: opt: 1329 z-score: 1411.0 E():
0, 33.6% identity in 1232 aa overlap or to AF040570.9
Amycolatopsis mediterranei rifamycin biosynthetic gene
cluster (3413 aa), 34.3% identity in 1162 aa overlap.
Also highly similar to mycobacterial putative polyketide
synthase: MTCY38_19, U00023.3 M. leprae,
MTCY22610.2, MTCY19H9_2, U00023.6 M. leprae, contains
PS00606 Beta-ketoacyl synthases active site."
/codon_start=1
/transl_table=1
/product="pks13"
/protein_id="CAA17864.1"
/db_xref="GI:2950419"
/db_xref="SPTREMBL:O53579"
/translation="MADVAESQENAPAEERALETVENKRWLMNMGKAVGAKPDSIDE
SVPAVELGSSRDVAVMAADIEDLTGLTSAVAFAPHTIESLATRIIEGEPETDLG
DDAEDMRTGPAERVDIAIVGLSTRPEGMPTPOTWALLLEGDGTLDLPDGGSEER
LEBPRLAARVAGARTBGYLRKDKIGFDSSEPAVAKTEADNIDPQRMALDELTEALEH
ARIPASLSRQAVGYIGSSNDYSFLAVSDPYAHRYAITGTSSITANRVSTFYDF
HGSRVITIDTACSSSLVAIHQGVQALRNGEADVAVAGVNALITPMVTLGDEIGAVLA
PDGRIKFSFADADGTYTSEGGGMVLKRVDDARVAGDAILIAGVSAVNHDRSGNLIT
APNDQADVILRAYKDGIDIPRTVDYIEAGTGTIIGDPIEALGLRVVGRGPADR
PALGAVKTVNGHLESAAASMAKVVIALQHDKLPSPINPAGSPYITDPDAMRLKMI

TTPTDMPRYGGYALAGVSPFGCGANAHVYREVLPRDYYEKEPEPEPEKAAEPAP
APTLGAALRDEDERGNTITSAVAKEEPEBELPGYTEERALRIKEALAEILAAOETAPL
VPLVASFUTSRKRAAAAEELADMQSPBGCASLSIEGRSISRNRHRSRAVYIAHH
DEAIKGLRAVAGKQAPNVFSVDSPTVITPVKREKHEMLPAHFGSDSTYMPGTHVA
IEKVDALVOBELGYSVLELIDDAQDYGIEETQVTFIPAIOALGELLRHGAKPAAYI
GOSIGEASAFEFAGLSLRDATRAICSRSHLMGSEAMLFGEYIRLMAIYEASDEIR
EVFSDPDLCEVYAAPTQYVIGGPPPOVDAILARAEKGFAKFAVTKGASHSOD
PLIGELTAELOGIXPTSPGCIESTVEGRYIKRGGPHIDHVEWKKGLRRSVFTHG
IRNAVDSGHTTFELARNPALMAYALITADAGLHDQLIPTLARKDEYSWSTVA
QLVYGHDLDIRLFSRASGPQDYANIPTPERFKREHMLPAHFGSDSTYMPGTHVA
PDGRHWEYAPRDGNVDLALVRAAAHVLPDQITAEQNAVAGDARIVTMTTRH
GGASQVHARIDESFTLVYDALVSRASSEVLPYVGAATAVAADGAPAPAEPTPAD
ADAEITLSDLTTRMPSGMTRWSPDSGETTAEERGLVGSAMGEPEPDLMEVYLIEL
GLDSIMAVRIKRVYEPDPLPPIOLTAVRANVYVKEILIEYAYEHRDEVQOJHEHOK
TQTAETARAQAEELHGRVGTKEPEYDEAGVAFPSONGEDPNTGALNADVPPRA
AERYETAWALVTGKSPGGLFENLPRLDDEAAKIAORLSRAGPTTAEDVLTSSNI
EALADKRVYLEAGQIDGFVFTLRARBAKGKVFVFNHSPAGSTVYVEPLGLRPLD
TPMGFEVERGESIEERAQOYVPKLIEMQGDGPLYVMSLGGVLAACAIGLRIGD
VRFVGLDAVAGEEIPOTKEIRKRWDRYAFAEKTENVTIPAIPEOLEELDECO
VRFVLDVAVSOSGQIPAGIIEHORTSYLDNRAIDTQIOPYDGHVTLVMDRYHDDAI
MEPRYIVROPDGGMGVEYSDLEVVPYIGGEHQAIDEPITAKGEHMSRALGQTEADR
TSEVGKO"
complement(4945..4948)
RBS
Query Match 31.3%; Score 31.6; DB 1; Length 23740;
Best Local Similarity 58.5%; Pred. No. 2.9e+02;
Matches 55; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

Search completed: November 6, 2002, 13:24:51
Job time : 568.625 secs


```

/organism="Brassica oleracea"
/strain="T01000DH3"
/db_xref="taxon:3712"
/clone="BOGA019"
/clone_1lb="BOGA"
/note="Vector: pHS01, Site_1: BstXI; 2-3 kb sheared
genomic DNA inserted into pHS01 using BstXI linkers"
BASE COUNT      149 a      246 c      238 g      165 t
ORIGIN

```

Query Match	36.2%	Score 36.6;	DB 17;	Length 798;
Best Local Similarity	65.1%	Pred. NO. 9.5;		
Matches 54; Conservative	0;	Mismatches 29;	Indels 0;	Gaps 0;

Qy 18 GCGACCCGGGACGGCTCGAGAGTGTGTCGATCCTTCGGCAAGGGGTGTCTGCC 77
Db 696 GCGCGAAGGCGACGCTCGGAGAAAGGCGCAAGCGCAACTCGCAGACGCAAGAGGCGTGTGCG 755
Qy 78 GTGAGTCAAGGTGGACGTGCGCC 100
Db 756 GCGCGTACAGAGTGGCTGTGCGCC 778

LOCUS	DEFINITION	BI780606/c	BI780606	387 bp	mRNA	linear	EST 23-JUL-2002
BI780606/c	EBES01_S0001_G23 R embryo sac, 4-6 DPA, no treatment, cv Optic, EBES01 Hordeum vulgare cDNA clone EBES01_S0001_G23 5', mRNA sequence.						

ACCESSION	BI780606	GI:21938718
VERSION	BI780606.2	
KEYWORDS	EST.	
SOURCE	Hordeum vulgare.	
ORGANISM	Hordeum vulgare	

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1	Triliciceae; Hordium.			
1	(bases 1 to 387)			
	Hedley, P., Liu, H., Caldwell, D., McCallum, N., Mudie, S., Cardie, L., Ramsay, L., Mchreay, G., Marshall, D.F.M. and Waugh, R.			
	Development of Bailey Transcriptome Resources			
	Unpublished (2001)			
	On Sep 26, 2001 this sequence version replaced g1:15783458.			

FEATURES
source
Contact: Waugh R, Marshall DF
Genome Dynamics/Computational Biology
Scottish Crop Research Institute
Invergowrie, Dundee, DD2 5DA, Scotland, UK
Tel: 00 44 1382 562731
Fax: 00 44 1382 562426
Email: eat@scri.sari.ac.uk
All sequence has a Phred quality score of 20 or over
Seq primer: M13 reverse.
Location/Qualifiers
1. .387
/organism="Hordeum vulgare"

```

/organism="Arabidopsis thaliana"
/cultivar="Optic"
/db_xref="taxon:4513"
/clone="Ebes01_SQ001_G23"
/clone_11b="embryo sac, 4-6 DPA, no treatment, cv Optic, Ebes01"
/tissue_type="embryo sac"
/dev_stage="4-6 DPA"
/lab_host="DH10B"
/notice="Vector: pSPORT1; Site_1: sal I; site_2: Not I; Non-normalised library, directionally cloned into pSPORT1. Derived from embryo sacs dissected from developing grains (4-6 days post anthesis) in glasshouse grown barley plants. Developed as part of the barley transcriptome resources of BBSRC/SEERAD funded cereal IGF (Investigating Gene Function) project."

```

	Query Match	Similarity	35.2%	Score 35.6	DB 13	Length 387
	Best Local	Similarity	60.2%	Pred. No. 15		
	Matches	59	Conservative	0	Mismatches	39
					Indels	0
					Gaps	0
Qy	4	GACGCTGCTGGCGGGCGAGCCGCGACGCGCTGCAGAGTGTCTGGCGATCTTGGCGCAA	63			
Db	291	GAGGGCGTTGGCGCTGAGTGCGGGCGGGCCCGATGACCGGGGTGCGCCGATCTGCCGGCCAT	232			
Qy	64	GGGGGTGTTCTGCCGTGCGTCAAGGTGAGAGTCGCCA	101			
Db	231	CTTGTTACACAGCCACCCCGTGAAGTGCAGCGACCCA	194			

RESULT 3	
BE740888/c	
LOCUS	979 bp mRNA linear EST 15-SEP-2000
DEFINITION	G015933ZP01 NIH_MGC_9 Homo sapiens cDNA clone IMAGE:3947188 5', mRNA sequence.
ACCESSION	BE740888

Accession	BE740880
Version	BE740888.1
Keywords	GI:10154880
Source	EST.
Organism	human.
	Homo sapiens

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:
Mammalia: Eutheria: Primates: Catarrhini: Hominoidea: Homo.
1 (bases 1 to 979)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC).
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: csapbs-remail.nih.gov

Tissue Procurement: DCM/D/PTP
 CDNA Library Preparation: Ling Hong/Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at: image.llnl.gov
 Plate: LCM808 row: h column: 05
 High quality sequence stop: 341.
 location/qualifiers

```

1. 979
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3947188"
/clone_1id="NH_MGC_9"
/tissue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: ovary; Vector: pOT8; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adapter: GGCACGAC(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
```

Query Match	34.5%	Score 34.8	DB 12	Length 979
Best Local Similarity	62.8%	Pred. No. 29		
Matches 54	Conservative 0	Mismatches 32	Indels 0	Gaps 0
QY	10	GCTGGCGGCGCAGCCGCGTCGCGAGAGTGTGCGCATCTTCGCGCAAGGGGGT	69	
Db	653	GGTGCGCGTGGCGCGCGTGGCTGCGTACCGAGAGGGGTGGGGTGGCGAGTGGCAGTGGGGGT	594	
QY	70	GTTCTGCGCTGAGTCAAGGTGACG	95	
Db	593	GGTGGCGCGTGGAGCGCAAGGGCGG	568	

RESULT 4
AM448711/c

LOCUS AM448711 678 bp mRNA linear EST 03-JAN-2001
 DEFINITION BR1180 BRY Triticum aestivum cDNA clone P26-1C, mRNA sequence.
 ACCESSION AM448711
 VERSION AM448711.1 GI:12019246
 KEYWORDS EST.
 SOURCE bread wheat.
 ORGANISM Triticum aestivum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidaeae ; Triticeae; Triticum.
 1 (bases 1 to 678)
 REFERENCE Clarke,B.C., Hobbs,M. and Appels,R.
 TITLE Genes active in developing wheat endosperm
 JOURNAL Unpublished (2000)
 COMMENT Contact: Bryan Clarke
 Division of Plant Industry
 C.S.I.R.O.
 GPO Box 1600, Canberra, ACT, Australia
 Tel: 61 2 6246 5054
 Fax: 61 2 6246 5000
 Email: bryan@pl.csiro.au.

FEATURES
 source
 1..678
 /organism="Triticum aestivum"
 /cultivar="Wynna"
 /db_xref="taxon:4565"
 /clone="P26-1C"
 /clone_lib="BRY"
 /cell_type="endosperm"

BASE COUNT 121 a 300 c 160 g 97 t
 ORIGIN

Query Match 33.5%; Score 33.8; DB 10; Length 678;
 Best Local Similarity 60.2%; Pred. No. 48;
 Matches 56; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 3 CGACGCTGCTGCGGCGCGCGCGCTCGACAGCTGCTGCGATCTTGGCGCA 62
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 133 CGACGAGTGTCCGACGACGAGGCTGCGGCGCTGCTGCGATCTTGGCGCA 74

QY 63 AGGGGGTGTCTGCCGTGAGTCAAGTGGACG 95
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 73 CGGGGAGAGTGTCCGCGCGGCGGCGGTGTCG 41

RESULT 5
 BE428772 401 bp mRNA linear EST 26-JUL-2000
 LOCUS MTD010.F10F990617 ITEC MTD Durum Wheat Root Library Triticum
 DEFINITION turgidum subsp. durum cDNA clone MTD010.F10, mRNA sequence.
 ACCESSION BE428772
 VERSION BE428772.1 GI:9426615
 KEYWORDS EST.
 SOURCE durum wheat.
 ORGANISM Triticum turgidum subsp. durum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidaeae ; Triticeae; Triticum.
 1 (bases 1 to 401)
 REFERENCE Anderson,O.A., Appels,R., Bailey,P., Blake,T., Close,T., Cloutier
 S., Dubcovsky,J., Feuillet,C., Gale,M., Graner,A., Gustafson,P.,
 Hermann,R.G., Holton,T., Jacquemin,J.M., Jia,J., Joudrier,P.,
 Langridge,P., Lazo,G.R., Lin,J.J., McGuire,P., Ogihara,T.,
 Pechioni,N., Qualset,C., Schuch,W., Selvaraj,G., Shariflou,M.,
 Sorrells,M., Warburton,M. and Wenzel,G.
 TITLE International Triticeae EST Cooperative (ITEC): Production of
 Expressed Sequence Tags for Species of the Triticeae
 JOURNAL Unpublished (2000)
 COMMENT Contact: Joudrier P
 INRA, Unite de Biochimie et Biologie Molculaire des Cereales
 2, place VIALA, 34060 Montpellier cedex 01 FRANCE
 Tel: 33 4 99 61 23 84
 Fax: 33 4 99 61 23 48

Email: joudrier@enscm.inra.fr
 International Triticeae EST Cooperative (ITEC)
 http://wheat.pw.usda.gov/genome.
 Location/Qualifiers
 1..401
 /organism="Triticum turgidum subsp. durum"
 /cultivar="Silliana"
 /db_xref="taxon:4567"
 /clone="MTD010.F10"
 /clone_lib="ITEC MTD Durum Wheat Root Library"
 /tissue_type="root"
 /dev_stage="3-day-old seedling, water-stressed"
 /note="Vector: pSPORT1; T7 primers used. See pSPORT1
 polylinker site. 0.3-2.0 kbp average insert size."

BASE COUNT 49 a 99 c 182 g 71 t
 ORIGIN

Query Match 33.3%; Score 33.6; DB 10; Length 401;
 Best Local Similarity 66.7%; Pred. No. 49;
 Matches 48; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 2 TCGACGCTGCTGCGGCGCGCGCGCTCGACAGTGTCTGCGATCTTGGCGCA 61
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 209 TCGACGCTGCTGCGGCGCGCGCTCGACAGTGTCTGCGATCTTGGCGCG 268

QY 62 AAGGGGTGTTC 73
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 269 GTGGTGTGCTTC 280

RESULT 6
 B1471502/c 534 bp mRNA linear EST 29-NOV-2001
 LOCUS sag21c02.y1 Gm-cl080 glycine max cDNA clone GENOME SYSTEMS CLONE
 DEFINITION ID: Gm-cl080-1659 5' similar to SW:5114.HUMAN Q15459 SPLICROSOE
 ASSOCIATED PROTEIN 114 ; mRNA sequence.
 ACCESSION B1471502
 VERSION B1471502.1 GI:15287611
 KEYWORDS EST.
 SOURCE soybean.
 ORGANISM Glycine max
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
 1 (bases 1 to 534)
 REFERENCE Shoemaker,R., Keim,P., Vodkin,L., Expelding,J., Coryell,V., Khanna
 A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
 Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
 Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
 R., Ratter,E., Kohn,S., Shin,T., Jackson,K., Cardenas,M., McCann
 R., Waterston,R. and Wilson,R.
 TITLE Public Soybean EST Project
 JOURNAL Unpublished (1999)
 COMMENT Contact: Shoemaker R/Public Soybean EST Project
 Public Soybean EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@wustl.wustl.edu
 This clone is available through: Reggen, Invitrogen Corp. 2130
 South Memorial Parkway Huntville, AL 35801 For further information
 call: (800)-533-4365 or contact via email: ccu@reggen.com
 High quality sequence stop: 422.
 Location/Qualifiers
 1..534
 /organism="Glycine max"
 /db_xref="taxon:3847"
 /clone="GENOME SYSTEMS CLONE ID: Gm-cl080-1659"
 /clone_lib="Gm-cl080"
 /tissue_type="Roots of 8 day old 'Bragg' supermodulating
 mutant NT5382 seedlings"

/dev-stage--"8 days old"
 /lab-host-"DH10B"
 /note-"Vector: plubscript II SK+, Site_1: EcoRI, Site_2:
 XhoI; The mRNA was isolated from roots of 8 day old
 'Bragg' supermodulating mutant MNS382 seedlings that were
 infected with Bradyrhizobium japonicum, strain USDA 110, 72
 hours prior to harvest. Dr. Gary Stacey generously donated
 the tissue. The roots were flash-frozen in liquid
 nitrogen. StrataGene's cDNA Synthesis Kit (catalog number
 200401) was used to synthesize the cDNA. First-strand
 synthesis was performed with 5-methyl dCTP, hence the
 ligated cDNA was hemimethylated. A modification of
 StrataGene's first-strand synthesis primer was used. An
 'anchor' nucleotide (V-A, C, or G) was added to the 3' end
 of the primer [GAGAGAGAGAGAGAGAGAGACTACTCTCGAG(T)18V] to
 anchor the primer at the 5' end of the poly(A) tract.
 After second-strand synthesis, the cDNA ends were filled
 in with cloned Pfu DNA, ligated to EcoRI adaptors and
 subsequently phosphorylated. The cDNA was then
 precipitated and redissolved in sterile, RNase-
 free water. The XhoI site within the first-strand
 synthesis primer was then restricted by digestion with
 XhoI from Promega (400U/ul); all XhoI sites in the cDNA
 would be protected by their hemimethylated status. The
 cDNA constructs were size-fractionated with a 500bp
 cutoff, using Sephacryl S-500 High Resolution (Pharmacia
 Biotech) in a 2-mm diameter column and a bed volume of
 approximately 1ml. The column eluent was precipitated,
 redissolved, and ligated into StrataGene's plubscript II
 XR predigested vector (plubscript II SK+) vector that
 has been digested with EcoRI and XhoI, and phosphorylated
 by StrataGene). This library was constructed in the
 laboratory of Dr. Paul Kelm and Dr. Virginia H. Coryell at
 Northern Arizona University."

Query Match Similarity	33.3%	Score 33.6	DB 13	Length 534
Best Local Similarity	70.3%	Pred. NO. 52		
Matches 45	Conservative 0	Mismatches 19	Indels 0	Gaps 0
QY	7	GGTCTGGCGGGGAGCCGCGAGCCCTGCGAGAGGTGCTGGCGATCTTGCGCAAGG 66		
Db	296	GGGAGTGGGGGAGAGACCTCTCTGTGTGGCACTGGTGGTGGCAGCTCTTGAGGCAAGGT 237		
QY	67	GGTG 70		
Db	236	GGTG 233		
RESULT 7				
LOCUS	BC839309			
DEFINITION	BC839309 700 bp mRNA linear EST 25-MAY-2001			
ACCESSION	Gm01_10f09_F Gm01_AAFG_ECONC_Glycine_max.cold.stressed.leaves			
VERSION	BC839309			
KEYWORDS	BC839309.1 GI:14205631			
SOURCE	soybean.			
ORGANISM	Glycine max			
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
AUTHORS	Singh,J.A., Robert,L.S., Lu,B., Zhu,L., De Moors,A., Couroux,P.,			
TITLE	Hartori,L.J., Hartori,J.I., Ouellet,T., Sprout,D. and Tinker,N.A.			
JOURNAL	Expressed Sequence Tags from Cold-Stressed Glycine max (soybean)			
COMMENT	leaves			
	Unpublished (2001)			
	Contact: Singh,J.A.			
	Eastern Cereal and Oilseed Research Centre			
	Agriculture and Agri-food Canada			

KC Neatby Bldg., Central Experimental Farm, Ottawa, Ontario, K1A
 0C6, Canada
 Tel: (613) 759-1662
 Fax: (613) 759-1701
 Email: slinghja@cm.agr.ca
 Location/Qualifiers
 1. .700
 FEATURES
 Source

```

/organism="Glycine max"
/cultivar="Maple Arrow"
/db_xref="taxon:3847"
/clone="Gm01_10f09"
/clone_lib="Gm01_AARC_ECORC_Glycine_max_cold_stressed_leaves"
/tissue_type="Leaves"

```

/dev_stage="15-day seedlings"
/node_vector: Bluescript SK-/XhoI-ECORI; Site 1: ECORI;
Site 2: XhoI; Plants were grown 12 days from seeds,
treated at 20c for 3 days, then leaves were collected.
Library was prepared using Unizap ZAP-cDNA synthesis kit
/ Packaged GigaPack III Gold."

Query Match	Best Local Similarity	Score	Pred.	No. Matches	Mismatches	Gaps
Matches 45;	Conservative	0;	Indels	0;	Gaps	0;

Oy 7 GGTCGTGGCGGAGCCGCCGACGCTGCAGAGTGTGGCGATCCTTGCGGCAAAAGG 66
|| ||| ||| ||| | | ||| ||| ||| ||| ||| ||| |||
Db 444 GGGATGTGGGGGAGGAGCCCTCTGTGTGGCAGTGTGTGGCACCTCTGAAGCAAAGT 503

Oy 67 GGTC GTG 70
||||
Db 504 GGTC GTG 507

RESULT 8
LOCUS BG839262 764 bp mRNA linear EST 25-MAY-2001
DEFINITION Gm01_10g08_F Gm01_AARC_ECORC_Glycine_max_cold_stressed_leaves
ACCESSION BG839262
VERSION BG839262.1 GI:14205584
KEYWORDS EST.
SOURCE soybean.
ORGANISM Glycine max
Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
Glycine.
REFERENCE 1 (bases 1 to 764)
AUTHORS Singh,J.A., Robert,L.S., Lu,B., Zhu,L., De Moors,A., Couroux,P.,
Harris,L.J., Hattori,U.I., Ouellet,T., Sprout,D. and Tinker,N.A.
TITLE Expressed Sequence Tags from Cold-Stressed Glycine max (soybean)
LEAVES
JOURNAL Unpublished (2001)
COMMENT Contact: Singh,J.A.
Eastern Cereal and Oilseed Research Centre
Agriculture and Agri-Food Canada
KW Neelby Bldg., Central Experimental Farm, Ottawa, Ontario, K1A
OC6, Canada
Tel: (613) 759-1662
Fax: (613) 759-1701
Email: singhja@em.agr.ca.
Location/Qualifiers
FEATURES
Source
1..764
/organism="Glycine max"
/cultivar="Maple Arrow"
/db_xref="taxon:3847"
/clone="Gm01_10g08"
/clone_lib="Gm01_AARC_ECORC_Glycine_max_cold_stressed_leaves"
/es

BASE COUNT
ORIGIN

```
210 a 152 c 203 g 186 t 13 others
      /dev_stage="15-day seedlings"
      /note="Vector: Bluescript SK-/XhoI-EcoRI; Site_1: EcoRI;
      Site_2: XhoI; Plants were grown 12 days from seeds,
      treated at 20C for 3 days, then leaves were collected.
      Library was prepared using Unilap ZAP-cDNA synthesis kit.
      /packaged gigaPack III Gold."
```

Query Match	33.3%	Score 33.6	DB 12	Length 764
Best Local Similarity	70.3%	Pred. No. 56		
Matches 45; Conservative	0	Mismatches 19	Indels 0	Gaps 0

DY 7 GGTGCTGGGGGGCGAGCCGGCAGCGCTTCGCAGAGTGTCTGGGATCCTTGCGGCAAAAGG 66
 || ||| | |||| | | |||| |||| |||| |||| ||||
Db 447 GGAGTGGGGGAGGACCTCTCTGTGTGGCAGTGTGTGGCACATCCTTGAGGCAAAGT 506

QY	67	GGTG	70
Db	507	GGTG	510

RESULT	9
LOCUS	BM816836
DEFINITION	BM816836 870 bp mRNA linear EST 05-MAR-2002 HB105A04.SK.ab1 HB Hordeum vulgare cDNA clone HB105A04.SK.ab1

ACCESSION	BM816836
VERSION	BM816836.1
KEYWORDS	GI:19152850
SOURCE	EST.
	<i>Hordeum vulgare</i> .

REFERENCE
Ozturk,N.Z., Michalowski,C.B., Brazille,S., Borchert,C., Palacio,C.
AUTHORS

TITLE	Monitoring large-scale changes in transcript abundance in drought- and salt-stressed barley
JOURNAL	Unpublished (2002)
COMMENT	Contact: Mark A. Fredrickson

University of Illinois
1201 W Gregory Dr, Urbana, IL 61801, USA
Tel: 2172655473
Email: bohnertlab@life.uiuc.edu

FEATURES
source location/Qualifiers
1. .870

```

/organism="Hordeum vulgare"
/strain="cv tokai"
/db_xref="taxon:4513"
/clone="HB105A04_SK.ab1"
/clone_1lb="HB"
/tissue_type="leaf"
/dev_stage="3 week old"
/note="6 and 10 hour drought stress by placing plants on
moist paper (75% rel. humidity) in light"
BASE COUNT      195 a      214 c      246 g      209 t      6 others
ORIGIN

```

	Query Match	33.3%	Score 33.6;	DB 14;	Length 870;
	Best Local Similarity	66.7%	Pred. No. 57;		
Matches	48:	Conservative	0;	Mismatches	24; Indels 0; Gaps 0.
OY	2	TGCAAGGTCTTGCCGGCGGACGCCGTCGCAGAGTCTTGCGCATCCTTGCGGCA	61		
db	86	TTGAGCGGTCTTGCCGGCGTCTGCATTAACCCGCTGAAGAGGTCCCGCTAAGCACGCTTGCGGCG	145		

QY	62	AAGGGGTGTC	73
Db	146	GTGGTGTGCTC	157

RESULT 10	LOCUS	DEFINITION
C67972/c	C67972	360 bp mRNA linear EST 22-SEP-1997
	C67972	yuji Kohara unpublished cDNA <i>Caenorhabditis elegans</i> CDNA clone yk308c8 5', mRNA sequence.

VERSION	C67972.1	GI:2426902
KEYWORDS	EST.	
SOURCE	Caenorhabditis elegans	
ORGANISM	Caenorhabditis elegans	

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae
; Rhabditidae; Pelodierinae; Caenorhabditis.
1 (bases 1 to 360)
Kohara,Y., Motolashv,T., Tabara,H., Watanabe,H., Sugimoto,A., Sano
,M., Miyata,A. and Nishigaki,A.
Expression map of the C.elegans genome
Unpublished (1996)
Contact: Yuji Kohara

National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 81-559-81-6854
Fax: 81-559-81-6855
Email: ykohara@lab.nig.ac.jp

```

/organism="Caenorhabditis elegans"
/strain="CB1489 him-8(e1489)"
/db-xref="taxon:6239"
/clocn="yk308c8"
/clocn_lib="yuj1 Kohara unpublished cDNA"
/sex="hermaphrodite, male"
/tissue_type="whole animal"
/dev_stage="varied"

```

Query Match	33.1%	Score 33.4	DB 14	Length 360
Best Local Similarity	68.7%	Pred. No. 54		
Matches 46	Conservative	0	Mismatches 21	Indels 0
				Gaps 0

```

27 7 GGTGCTGGCGGCGAGCCGGCAGCGCTCGAGAGTGCTGGCATCTTGGGGCAAAGG 66
    ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 149 GGTGGGGGCGGCGCAGCGCGGCATACCGACGAGAGTGGGGCTCTCCCATCATGTT 90

```

Qy	67	GGTGTTC	73
Db	89	GGTGTTC	83

LOCUS	DEFINITION	VERSION	KEYWORDS	SOURCE
RESULT 11 Btj31405	Btj31405 unpublished oligo-capped cDNA library C. elegans L1 stages			
	617 bp			
	linear			
	EST 23-JAN-2002			
	Caenorhabditis elegans cDNA clone YK1054d07 3', mRNA sequence.			
	Btj31405			
	Btj31405			
	Btj31405.1 GI:18291562			
	EST.			
	Caenorhabditis elegans.			

ORGANISM	<i>Caenorhabditis elegans</i> Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae ; Rhabditidae; Pelodertinae; Caenorhabditis. 1 (bases 1 to 617)
REFERENCE	Kohara,Y., Shin-I,T., Thierry-Mieg,J., Thierry-Mieg,D., Suzuki,Y. and Sugano,S. A complementary view of the C.elegans genome Unpublished (2002)
TITLE	Unpublished (2002)
JOURNAL	Contact: Tadasu Shin-I
COMMENT	

Query Match	30.7%	Score 31;	DB 10;	Length 969;
Best Local Similarity	72.7%	Pred. NO. 0.84;		
Matches	40;	Conservative	0;	Mismatches 15;
			Indels	0;
			Gaps	0;

Oy 13 GCGGGCCAGCCGCGCTGCAGAGGTGCTGGCCATCCTTCGGGCCAAAGGG 67
| | | | | | | | | | | | | | | |
Db 513 GCCGGGTTCGCGCGCGATCGACGCGTGGCGCAGCCCTGCCACCAAGGGG 567

RESULT 2

US-09-747-810-1/c
; Sequence 1, Application US/09747810
; Patent No. US20020012903A1

```

1  APPLICANT: Goldman, Steven A.
2  APPLICANT: Okano, Hideyuki
3  TITLE OF INVENTION: A METHOD FOR ISOLATING AND PURIFYING MULTIPOTENTIAL NEURAL PROGEN
4  TITLE OF INVENTION: CELLS AND MULTIPOTENTIAL NEURAL PROGENITOR CELLS
5  FILE REFERENCE: 19603/3580
6  CURRENT APPLICATION NUMBER: US/09/747,810
7  CURRENT FILING DATE: 2000-12-22
8  PRIOR APPLICATION NUMBER: 60/173,003
9  PRIOR FILING DATE: 1999-12-23
10 NUMBER OF SEQ ID NOS: 2
11
12 SEQ ID NO 1
13 LENGTH: 52216
14 TYPE: DNA
15 ORGANISM: Homo sapiens
16
17 US-09-747-810-1

```

Query Match	29.1%;	Score 29.4;	DB 10;	Length 52216;
Best Local Similarity	66.7%;	Pred. NO. 4;		
Matches 42;	Conservative	0;	Mismatches 21;	Indels 0;
				Gaps 0;

Oy 6 CGGTCTGGCGGCGAGCCGGCAGCGCTGCAGAGGTGCTGGCGATCTTGGGCCAAAG 65
 ||| | |||| | |||| ||||| ||| ||| |||
Db 52183 CCGCGGCGGCGCGGCGGCGGCGCTGCGGCGGCGGCAATGAGGAGCGCGCGAAG 52124

QY	66	GGG	68
Db	52123	GGG	52121

RESULT 3

US-09-822-849A-22
; Sequence 22, Application US/09822849A
; Patent No. US20020045170A1

```

? APPLICANT: Wong, Gordon G.
? APPLICANT: Clark, Hilary
? APPLICANT: Fechtel, Kim
? APPLICANT: Agostino, Michael J.
? APPLICANT: Howes, Steven H.
? APPLICANT: Resnick, Richard J.
? APPLICANT: Gulikota, Kamalakar
? APPLICANT: Graham, James R.
? APPLICANT: Genetics Institute, Inc.
? TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
? FILE REFERENCE: GIN 6403
? CURRENT APPLICATION NUMBER: US/09/822,849A
? CURRENT FILING DATE: 2001-09-04
? PRIOR APPLICATION NUMBER: 60/195,582
? PRIOR FILING DATE: 2000-04-06
? NUMBER OF SEQ ID NOS: 598
? SOFTWARE: PatentIn Ver. 2.0
? SEQ ID NO 22
? LENGTH: 1204
? TYPE: DNA
? ORGANISM: Homo sapiens
? US-09-822-849A-22

```

Query match 28.78; Score 29; DB 10; Length 1204;

Best Local Similarity 58.8%; Pred. No. 3.3;
Matches 50; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

Oy 9 TGCATGCGGGCCGAGCCGGCAGAGGCTCGCAGAGTGTGGCGATCTTTGCCGCAAAAGGGGG 68
 || | | | | | | | | | | | | | | |
Db 531 TGATGACGCGTGAGCTTTGAGCGGCTGGCTGMAACGCGTGGCCACAACTCGGGGCAGGCAAG 590

Oy	69	TGTTCTGCCGTCGAGTCAAGCTGGA	93
Db	591	GGTTTCTCGCCCGTGTGTACGCGGA	615

RESULT 4

US-09-964-824A-252/c
; Sequence 252, Application US/09964824A
; Patent No. US20020102531A1
; CURRENT INFORMATION

```

APPLICANT: Horriagan, Stephen
TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Sigm
TITLE OF INVENTION: Sets
FILE REFERENCE: 689230-73
CURRENT APPLICATION NUMBER: US/09/964,824A
CURRENT FILING DATE: 2001-09-27
PRIOR APPLICATION NUMBER: US/60/236,033
PRIOR FILING DATE: 2000-09-28
PRIOR APPLICATION NUMBER: US/60/236,032
PRIOR FILING DATE: 2000-09-28
PRIOR APPLICATION NUMBER: US/60/236,028
PRIOR FILING DATE: 2000-09-28
NUMBER OF SEQ ID NOS: 583
SOFTWARE: PatentIn version 3.0
SEQ ID NO 252
LENGTH: 2678
TYPE: DNA
ORGANISM: Homo sapiens
US-09-964-824A-252

```

query Match	28.7%;	Score 29;	DB 10;	Length 2678;
Best Local Similarity	58.8%;	Pred. No. 3.6;		
Matches 50;	Conservative 0;	Mismatches 35;	Indels 0;	Gaps 0;

[illegible]

Oy 69 TGTTCGCGGTCGAGTCAAGGTGA 93
 ||| | ||| | |||
 Db 2130 GGTTCGCGCGGTCGTACCGCGA 2106

RESULT 5

US-09-861-893-42
; Sequence 42, Application US/09861893
; Patent No. US20020045257A1
GENERAL INFORMATION

```

APPLICANT: Feinberg, Andrew
APPLICANT: Scritchman-Altmashanu, Liora
APPLICANT: Jiang, Shan
TITLE OF INVENTION: METHODS FOR ASSAYING GENE IMPRINTING AND
TITLE OF INVENTION: METHYLATED CGP ISLANDS
FILE REFERENCE: 01107.00128
CURRENT APPLICATION NUMBER: US/09/861,893
CURRENT FILING DATE: 2001-05-22
PRIOR APPLICATION NUMBER: 60/206,158
PRIOR FILING DATE: 2000-05-22
PRIOR APPLICATION NUMBER: 60/206,161
PRIOR FILING DATE: 2000-05-22
NUMBER OF SEQ ID NOS: 77
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 42
LENGTH: 1578

```

ORGANISM: *Homo sapiens*


```
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(1578)
; OTHER INFORMATION: n = A,T,C or G
US-09-861-893-42
```

```
Query Match          28.5%; Score 28.8; DB 10; Length 1578;
Best Local Similarity 57.3%; Pred. No. 3.9;
Matches 51; Conservative 0; Mismatches 38; Indels 0; Gaps 0;
```

```
Oy 4 GACCGTGTGCGGGCCGAGCCGCTGCAGAGTGTGCGGATCTTCCGGCAA 63
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 56 GNCCTGTCTCGGGCCGCGCCGAGCCGAGTACAGCTCCGCCAAGACTGTCTCT 115
```

```
Oy 64 GGGGGTGTTCGCGCTGCGATCAAGGTGG 92
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 116 GTGGCTTTCTTCTTCTTAAGTCTGGAGG 144
```

RESULT 6

```
US-09-784-962-1
; Sequence 1, Application US/09784962
; Patent No. US20020146431A1
```

```
; GENERAL INFORMATION:
; APPLICANT: ALLAN, Gordon M.
; APPLICANT: MEEHAN, Brian M.
; APPLICANT: ELLIS, John A.
; APPLICANT: KRAKOWKA, George S.
; APPLICANT: AUDONNET, Jean-Christophe F.
; TITLE OF INVENTION: PORCINE CIRCOVIRUS AND PARVOVIRUS VACCINE
; FILE REFERENCE: 454313-2338
; CURRENT APPLICATION NUMBER: US/09/784,962
; PRIOR FILING DATE: 2001-02-16
; PRIOR APPLICATION NUMBER: 09/347,594
; PRIOR FILING DATE: 1999-07-04
; PRIOR APPLICATION NUMBER: 98 08777
; PRIOR FILING DATE: 1998-07-06
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.1
```

```
; SEQ ID NO 1
; LENGTH: 1767
; TYPE: DNA
; ORGANISM: Porcine circovirus
US-09-784-962-1
```

```
Query Match          28.5%; Score 28.8; DB 10; Length 1767;
Best Local Similarity 58.0%; Pred. No. 3.9;
Matches 51; Conservative 0; Mismatches 37; Indels 0; Gaps 0;
```

```
Oy 1 ATCGACGGTGTGCGGGCCGAGCCGCTGCAGAGTGTGCGGATCTTCCGGC 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 209 AACGGTGGCGGGGTGGACGAGCCAGGCGCGCGGAGATCTGGCCAAAGATGGCTTC 268
```

```
Oy 61 AAAGGGGGTGTTCGCGCTGCGATCAAG 88
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 269 GGGGGCGGTGTCTTCTTCTCCGTAACG 296
```

```
US-09-784-962-2
; Sequence 2, Application US/09784962
; Patent No. US20020146431A1
; GENERAL INFORMATION:
```

```
; APPLICANT: ALLAN, Gordon M.
; APPLICANT: MEEHAN, Brian M.
; APPLICANT: ELLIS, John A.
; APPLICANT: KRAKOWKA, George S.
; APPLICANT: AUDONNET, Jean-Christophe F.
; TITLE OF INVENTION: PORCINE CIRCOVIRUS AND PARVOVIRUS VACCINE
; FILE REFERENCE: 454313-2338
; CURRENT APPLICATION NUMBER: US/09/784,962
; CURRENT FILING DATE: 2001-02-16
; PRIOR APPLICATION NUMBER: 09/347,594
```

```
RESULT 7
US-09-784-962-2
; Sequence 2, Application US/09784962
; Patent No. US20020146431A1
; GENERAL INFORMATION:
```

```
; APPLICANT: ALLAN, Gordon M.
; APPLICANT: MEEHAN, Brian M.
; APPLICANT: ELLIS, John A.
; APPLICANT: KRAKOWKA, George S.
; APPLICANT: AUDONNET, Jean-Christophe F.
; TITLE OF INVENTION: PORCINE CIRCOVIRUS AND PARVOVIRUS VACCINE
; FILE REFERENCE: 454313-2338
; CURRENT APPLICATION NUMBER: US/09/784,962
; CURRENT FILING DATE: 2001-02-16
; PRIOR APPLICATION NUMBER: 09/347,594
```

```
; PRIOR FILING DATE: 1999-07-04
; PRIOR APPLICATION NUMBER: 98 08777
; PRIOR FILING DATE: 1998-07-06
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1767
; TYPE: DNA
; ORGANISM: Porcine circovirus
US-09-784-962-2
```

```
Query Match          28.5%; Score 28.8; DB 10; Length 1767;
Best Local Similarity 58.0%; Pred. No. 3.9;
Matches 51; Conservative 0; Mismatches 37; Indels 0; Gaps 0;
```

```
Oy 1 ATCGACGGTGTGCGGGCCGAGCCGCTGCAGAGTGTGCGGATCTTCCGGC 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 209 AACGGTGGCGGGGTGGACGAGCCAGGCGCGCGGAGATCTGGCCAAAGATGGCTTC 268
```

```
Oy 61 AAAGGGGGTGTTCGCGCTGCGATCAAG 88
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 269 GGGGGCGGTGTCTTCTTCTCCGTAACG 296
```

```
US-09-884-514-1
; Sequence 1, Application US/09884514
; Patent No. US20020146432A1
```

```
; GENERAL INFORMATION:
; APPLICANT: ALLAN, Gordon
; APPLICANT: MEEHAN, Brian
; APPLICANT: CLARK, Edward
; APPLICANT: HAINES, Deborah
; APPLICANT: HASSARD, Lori
; APPLICANT: HARDING, John
; APPLICANT: CHARREYRE, Catherine E.
; APPLICANT: CHAPUIS, Gilles E.
; APPLICANT: NEWTONARDS, Francis Mcneilly
; TITLE OF INVENTION: NEW PORCINE CIRCOVIRUSES, VACCINES AND DIAGNOSTIC
; TITLE OF INVENTION: REAGENTS
```

```
; FILE REFERENCE: ALLAN
; CURRENT APPLICATION NUMBER: US/09/884,514
; CURRENT FILING DATE: 2000-06-19
; PRIOR APPLICATION NUMBER: FR 9800873
; PRIOR FILING DATE: 1998-01-22
; PRIOR APPLICATION NUMBER: FR 9803707
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: FR 97/12382
; PRIOR FILING DATE: 1997-10-03
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.0
```

```
; SEQ ID NO 1
; LENGTH: 1767
; TYPE: DNA
; ORGANISM: Porcine circovirus
US-09-884-514-1
```

```
Query Match          28.5%; Score 28.8; DB 10; Length 1767;
Best Local Similarity 58.0%; Pred. No. 3.9;
Matches 51; Conservative 0; Mismatches 37; Indels 0; Gaps 0;
```

```
Oy 1 ATCGACGGTGTGCGGGCCGAGCCGCTGCAGAGTGTGCGGATCTTCCGGC 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 209 AACGGTGGCGGGGTGGACGAGCCAGGCGCGCGGAGATCTGGCCAAAGATGGCTTC 268
```

```
Oy 61 AAAGGGGGTGTTCGCGCTGCGATCAAG 88
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 269 GGGGGCGGTGTCTTCTTCTCCGTAACG 296
```

```
US-09-884-514-2
; Sequence 2, Application US/09884514
```

```
RESULT 9
US-09-884-514-2
; Sequence 2, Application US/09884514
```

```
; APPLICANT: ALLAN, Gordon M.
; APPLICANT: MEEHAN, Brian M.
; APPLICANT: ELLIS, John A.
; APPLICANT: KRAKOWKA, George S.
; APPLICANT: AUDONNET, Jean-Christophe F.
; TITLE OF INVENTION: PORCINE CIRCOVIRUS AND PARVOVIRUS VACCINE
; FILE REFERENCE: 454313-2338
; CURRENT APPLICATION NUMBER: US/09/784,962
; CURRENT FILING DATE: 2001-02-16
; PRIOR APPLICATION NUMBER: 09/347,594
```

```
; Patent No. US20020146432A1
; GENERAL INFORMATION:
; APPLICANT: ALLAN, Gordon
; APPLICANT: MEEHAN, Brian
; APPLICANT: CLARK, Edward
; APPLICANT: HAINES, Deborah
; APPLICANT: HASSARD, Lori
; APPLICANT: HARDING, John
; APPLICANT: CHAREYRE, Catherine E.
; APPLICANT: CHAPUIS, Gilles E.
; APPLICANT: NEWMONADS, Francis McNelly
; TITLE OF INVENTION: NEW PORCINE CIRCOVIRUSES, VACCINES AND DIAGNOSTIC
; FILE REFERENCE: REAGENTS
; CURRENT APPLICATION NUMBER: US/09/884,514
; CURRENT FILING DATE: 2000-06-19
; PRIOR APPLICATION NUMBER: FR 9800873
; PRIOR FILING DATE: 1998-01-22
; PRIOR APPLICATION NUMBER: FR 9803707
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: FR 97/12382
; PRIOR FILING DATE: 1997-10-03
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 2
; LENGTH: 1767
; TYPE: DNA
; ORGANISM: Porcine circovirus
; US-09-884-514-2
```

```
Query Match      28.5%; Score 28.8; DB 10; Length 1767;
Best Local Similarity 58.0%; Pred. No. 3.9;
Matches 51; Conservative 0; Mismatches 37; Indels 0; Gaps 0;
```

```
QY 1 ATGAGCGTCTGCGCGCGCGAGCGCTCGCAGAGTCTGCGCATCTTGGCGC 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 209 AACGGTGGCGGGGTGACGAGCGCGCGCGAGAGTCTGCGCAAGATGGCTGC 268
QY 61 AAAGGGGTGTCTGCGCTCGAGTCAAG 88
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 269 GGGGGCGGTCTCTTCTTCGCGGTACG 296
```

```
RESULT 10
US-09-917-800A-1636/C
; Sequence 1636, Application US/09917800A
; Patent No. US20020119462A1
; GENERAL INFORMATION:
; APPLICANT: Mendrick, Donna
; APPLICANT: Porter, Mark
; APPLICANT: Johnson, Kory
; APPLICANT: Castle, Arthur
; APPLICANT: Elashoff, Michael
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Molecular Toxicology Modelling
; FILE REFERENCE: 44921-5038-US
; CURRENT APPLICATION NUMBER: US/09/917,800A
; CURRENT FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: US 60/222,040
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: US 60/222,880
; PRIOR FILING DATE: 2000-11-02
; PRIOR APPLICATION NUMBER: US 60/290,029
; PRIOR FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: US 60/290,645
; PRIOR FILING DATE: 2001-05-15
; PRIOR APPLICATION NUMBER: US 60/292,336
; PRIOR FILING DATE: 2001-05-22
; PRIOR APPLICATION NUMBER: US 60/295,798
; PRIOR FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: US 60/297,457
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,884
```

```
; PRIOR FILING DATE: 2001-06-19
; PRIOR APPLICATION NUMBER: US 60/303,459
; PRIOR FILING DATE: 2001-07-09
; NUMBER OF SEQ ID NOS: 1740
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 1636
; LENGTH: 2540
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020119462A1 NM_017127
; US-09-917-800A-1636
```

```
Query Match      28.3%; Score 28.6; DB 10; Length 2540;
Best Local Similarity 59.0%; Pred. No. 4.7;
Matches 49; Conservative 0; Mismatches 34; Indels 0; Gaps 0;
```

```
QY 1 ATGAGCGTCTGCGCGCGAGCGCTCGCAGAGTCTTGGCATCTTGGCGC 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 479 AGCGGCGGTATGCGCGCGCGAGCGCGCGCGGTGTGCGCGCGCGCGCG 420
QY 61 AAAGGGGTGTCTGCGCTCGAGTCAAG 83
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 419 AGAGTTGGAGCCGCCACCAAG 397
```

```
RESULT 11
US-09-832-320-3
; Sequence 3, Application US/09832320
; Patent No. US20010049834A1
; GENERAL INFORMATION:
; APPLICANT: Crane, Edmund H.
; APPLICANT: Crane, Virginia C.
; TITLE OF INVENTION: Polynucleotide and Methods of Use
; FILE REFERENCE: 35718/214291
; CURRENT APPLICATION NUMBER: US/09/832,320
; CURRENT FILING DATE: 2001-04-10
; PRIOR APPLICATION NUMBER: US 60/195,801
; PRIOR FILING DATE: 2000-04-10
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 612
; TYPE: DNA
; ORGANISM: zea mays
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(612)
; US-09-832-320-3
```

```
Query Match      27.9%; Score 28.2; DB 10; Length 612;
Best Local Similarity 59.3%; Pred. No. 5.2;
Matches 48; Conservative 0; Mismatches 33; Indels 0; Gaps 0;
```

```
QY 7 GGTGCTGGCGCGAGCGCGAGCGCTCGCAGAGTCTGCGCATCTTGGCGCAAGGG 66
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 369 GGGGTGGCGAGCTACCGCGCGCGCGCGCGCGAGGTGTGCGGTGGCGGAGGG 428
QY 67 GGTGCTTCTGCGCTCGAGTCAA 87
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 429 GCGGTACTACACCCACGCCAA 449
```

```
RESULT 12
US-09-832-320-1
; Sequence 1, Application US/09832320
; Patent No. US20010049834A1
; GENERAL INFORMATION:
; APPLICANT: Crane, Edmund H.
; APPLICANT: Crane, Virginia C.
; TITLE OF INVENTION: Maize Pathogenesis-Related
; TITLE OF INVENTION: Polynucleotide and Methods of Use
```

FILE REFERENCE: 35718/214291
CURRENT APPLICATION NUMBER: US/09/832,320
CURRENT FILING DATE: 2001-04-10
PRIOR APPLICATION NUMBER: US 60/195,801
PRIOR FILING DATE: 2000-04-10
NUMBER OF SEQ ID NOS: 3
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 898
TYPE: DNA
ORGANISM: Zee mays
FEATURE:
NAME/KEY: CDS
LOCATION: (63)...(674)
US-09-832-320-1

Query Match 27.9%; Score 28.2; DB 10; Length 898;
Best Local Similarity 59.3%; Pred. No. 5.4;
Matches 48; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

Qy 7 GGTGCTGGCGGCGGCGGCGGCTGCTGCGGATCCTTGGCGCAAGG 66
Db 431 GGGTGGGCGAGCTACCCGCGCGCCGCCGAGGTGGCTGTGGCGGAGG 490
Qy 67 GGTGTTCTGCCGTCGAGTCAA 87
Db 491 GCGGTACTACACCCAGCCAA 511

RESULT 13
US-09-864-761-12032
Sequence 12032, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
FILE REFERENCE: Aeomica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263,6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
SEQ ID NO 12032
LENGTH: 535
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC010905.3
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.1
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.4
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.1
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.3
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 5.2
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.2
US-09-864-761-12032

Query Match 27.7%; Score 28; DB 10; Length 535;
Best Local Similarity 58.3%; Pred. No. 5.8;
Matches 49; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

Qy 14 GCGGCGAGCCGCGGCGGCGGCTGCTGCGGATCCTTGGCGCAAGGSGTGTG 73
Db 419 GCGGGGCTGTGTGCGCGCGCTGCGAGCAGTGTGGTTCAGTGCGGCGGCTTGGTG 478
Qy 74 TGCCGTCGAGTCAGGTGAGCTG 97
Db 479 GGCCTGCACCTAGGAGGCCGCGC 502

RESULT 14
US-09-834-975-775
Sequence 775, Application US/09834975
Patent No. US20020110815A1
GENERAL INFORMATION:
APPLICANT: Lillie, James
APPLICANT: Brown, Jeffrey
APPLICANT: Bolt, Andrew
APPLICANT: Van Hufel, Christophe
TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS AND METHODS
TITLE OF INVENTION: FOR THE IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
FILE REFERENCE: MRI-016B
CURRENT APPLICATION NUMBER: US/09/834,975
CURRENT FILING DATE: 2001-04-13
PRIOR APPLICATION NUMBER: 60/197,538
PRIOR FILING DATE: 2000-04-14
NUMBER OF SEQ ID NOS: 1046
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 775
LENGTH: 3376
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc.feature
LOCATION: (1)...(3376)
OTHER INFORMATION: n = A,T,C or G
US-09-834-975-775

Query Match 27.7%; Score 28; DB 10; Length 3376;
Best Local Similarity 63.2%; Pred. No. 7.3;
Matches 43; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

Qy 3 CGACGCTGCTGGCGGCGGCGGCGGCTGCTGCGGATCCTTGGCGCAAG 62
Db 65 GCGGCGCGCGGCGGCACTGAGGCGGCGGCGGCTGCGGCGGCGGCTTACCGCGGCGG 124
Qy 63 AGGGGCTG 70

DB 125 ACCTGGTG 132

RESULT 15
US-09-935-428A-24

; Sequence 24, Application US/09935428A

; Patent No. US2002010639A1

; GENERAL INFORMATION:

; APPLICANT: WANG, LI

; APPLICANT: BABIUK, LORNE A.

; APPLICANT: POTTER, ANDREW A.

; APPLICANT: WILSON, PHILIP

; TITLE OF INVENTION: POSTWEANING MULTISYSTEM WASTING SYNDROME VIRUS FROM

; TITLE OF INVENTION: PIGS

; FILE REFERENCE: 9000-0040

; CURRENT APPLICATION NUMBER: US/09/935,428A

; CURRENT FILING DATE: 2001-08-20

; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/209,961

; PRIOR FILING DATE: EARLIER FILING DATE: 1998-12-10

; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/069,233

; PRIOR FILING DATE: EARLIER FILING DATE: 1997-12-11

; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/069,750

; PRIOR FILING DATE: EARLIER FILING DATE: 1997-12-16

; NUMBER OF SEQ ID NOS: 24

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 24

; LENGTH: 1343

; TYPE: DNA

; ORGANISM: Porcine Circovirus Type II

; US-09-935-428A-24

Query Match 27.5%; Score 27.8; DB 10; Length 1343;

Best Local Similarity 62.0%; Pred. No. 7.4;

Matches 44; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

QY 1 ATGACGGTGTCTGGCGGCGAGCGGCTGCGACAGAGTGTGGCGATCCTTGGGGC 60

DB 1205 ACCTGGTGCGGCGGCGAGCGGCGGCGGCGGCGGCGGAGATGTGCGCAAGATGGCTGC 1264

QY 61 AAGGGGGTGT 71

DB 1265 GGGGGCGGTGT 1275

Search completed: November 6, 2002, 20:31:29
Job time : 32.125 secs

; Sequence 1, Application US/09335409


```
RESULT 6
US-09-568-486-1
; Sequence 1, Application US/09568486
; Patent No. 6355459
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,486
; PRIOR FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-486-1

Query Match          99.0%; Score 100; DB 4; Length 68750;
Best Local Similarity 100.0%; Pred. No. 2.9e-19;
Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 TCGACGGTGTCTGCGGGGCGGACCGCGACGCTCGCAGAGTGTCGCGATCTTCGGCGCA 61
Db 39113 TCGACGGTGTCTGCGGGGCGGACCGCGACGCTCGCAGAGTGTCGCGATCTTCGGCGCA 39172

OY 62 AAGGGGGTGTCTGCGCGTGCAGTCAAGGTGAGCGTCGCCA 101
Db 39173 AAGGGGGTGTCTGCGCGTGCAGTCAAGGTGAGCGTCGCCA 39212

RESULT 7
US-09-568-472-1
; Sequence 1, Application US/09568472
; Patent No. 6358719
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,472
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-472-1

Query Match          99.0%; Score 100; DB 4; Length 68750;
Best Local Similarity 100.0%; Pred. No. 2.9e-19;
Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 TCGACGGTGTCTGCGGGGCGGACCGCGACGCTCGCAGAGTGTCGCGATCTTCGGCGCA 61
Db 39113 TCGACGGTGTCTGCGGGGCGGACCGCGACGCTCGCAGAGTGTCGCGATCTTCGGCGCA 39172

OY 62 AAGGGGGTGTCTGCGCGTGCAGTCAAGGTGAGCGTCGCCA 101
Db 39173 AAGGGGGTGTCTGCGCGTGCAGTCAAGGTGAGCGTCGCCA 101
```

```
Db 39173 AAGGGGGTGTCTGCGCGTGCAGTCAAGGTGAGCGTCGCCA 39212

RESULT 8
US-09-567-899-1
; Sequence 1, Application US/09567899
; Patent No. 6383787
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/567,899
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-567-899-1

Query Match          99.0%; Score 100; DB 4; Length 68750;
Best Local Similarity 100.0%; Pred. No. 2.9e-19;
Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 TCGACGGTGTCTGCGGGGCGGACCGCGACGCTCGCAGAGTGTCGCGATCTTCGGCGCA 61
Db 39113 TCGACGGTGTCTGCGGGGCGGACCGCGACGCTCGCAGAGTGTCGCGATCTTCGGCGCA 39172

OY 62 AAGGGGGTGTCTGCGCGTGCAGTCAAGGTGAGCGTCGCCA 101
Db 39173 AAGGGGGTGTCTGCGCGTGCAGTCAAGGTGAGCGTCGCCA 39212

RESULT 9
US-08-764-233A-1
; Sequence 1, Application US/08764233A
; Patent No. 5716849
; GENERAL INFORMATION:
; APPLICANT: Ligon, James M.
; APPLICANT: Schupp, Thomas
; APPLICANT: Beck, James J.
; APPLICANT: Hill, Dwight S.
; APPLICANT: Neff, Suzanna
; APPLICANT: Ryals, John A.
; TITLE OF INVENTION: Genes For The Biosynthesis Of Soraphen
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 520 White Plains Road, P.O. Box 2005
; CITY: Tarrytown
; STATE: NY
; COUNTRY: USA
; ZIP: 10591
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/764,233A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/729,214
```

FILING DATE: 09-OCT-1996
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/258,261
 FILING DATE: 08-JUN-1994
 ATTORNEY/AGENT INFORMATION:
 NAME: Melgs, J. Timothy
 REGISTRATION NUMBER: 38,241
 REFERENCE/DOCKET NUMBER: 1506/CIP6
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (919) 541-8587
 TELEFAX: (919) 541-8689
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 49377 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 ORIGINAL SOURCE:
 ORGANISM: Sorangium cellulosum
 IMMEDIATE SOURCE:
 CLONE: p98/1, pUL3, and pVKM15
 FEATURE:
 NAME/KEY: misc_feature
 LOCATION: 383..760
 OTHER INFORMATION: /product= "SorR"
 OTHER INFORMATION: /note= "This gene encodes a protein that is highly homologous to type I PKSS from *Saccharopolyspora erythraea*."
 OTHER INFORMATION: the reductase domains of type I PKSS such as eryA from
 FEATURE:
 NAME/KEY: misc_feature
 LOCATION: 927..19874
 OTHER INFORMATION: /product= "SorA"
 OTHER INFORMATION: /note= "Gene product is highly homologous to type I PKSS that
 OTHER INFORMATION: are known to be involved in the synthesis of polyketide
 OTHER INFORMATION: compounds."
 FEATURE:
 NAME/KEY: misc_feature
 LOCATION: 942..7115
 OTHER INFORMATION: /product= "Module 1 of SorA"
 FEATURE:
 NAME/KEY: misc_feature
 LOCATION: 7203..12884
 OTHER INFORMATION: /product= "Module 2 of SorA"
 FEATURE:
 NAME/KEY: misc_feature
 LOCATION: 13455..19616
 OTHER INFORMATION: /product= "Module 3 of SorA"
 FEATURE:
 NAME/KEY: misc_feature
 LOCATION: 19671..46318
 OTHER INFORMATION: /product= "SorB"
 OTHER INFORMATION: /note= "Gene product is highly homologous to type I PKS genes
 FEATURE:
 NAME/KEY: misc_feature
 LOCATION: 19870..24556
 OTHER INFORMATION: /product= "Module 1 of SorB"
 FEATURE:
 NAME/KEY: misc_feature
 LOCATION: 24638..30820
 OTHER INFORMATION: /product= "Module 2 of SorB"
 FEATURE:
 NAME/KEY: misc_feature
 LOCATION: 30881..35446
 OTHER INFORMATION: /product= "Module 3 of SorB"
 FEATURE:
 NAME/KEY: misc_feature
 LOCATION: 35528..40114
 OTHER INFORMATION: /product= "Module 4 of SorB"
 FEATURE:
 NAME/KEY: misc_feature
 LOCATION: 40190..46318
 OTHER INFORMATION: /product= "Module 5 of SorB"

```

; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 46851..47891
; OTHER INFORMATION: /product="Sorm"
; OTHER INFORMATION: /note="The protein encoded by the sorm gene is highly
; OTHER INFORMATION: homologous to the methyltransferase from Streptomyces
; OTHER INFORMATION: hygroscopicus that is involved in the synthesis of the
; OTHER INFORMATION: polyketide rapamycin."
US-08-764-233A-1

Query Match          52.1%; Score 52.6; DB 1; Length 49377;
Best Local Similarity 70.7%; Pred. No. 3.6e-06;
Matches 70; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

OY 2 TCGACGCTGCTGGCGGCGGACCGCGACGGCTTCGACAGAGTCTCTGCGATCCTTGCGGCA 61
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3300 TCAACGGTGTCTGGCGGCGGACCGCGACGGCTTCGACAGAGTCTCTGCGATCCTTGCGGCG 3359

OY 62 AAGGGGGTGTCTGCGCTGCAGTCAAGGTGAGACGTCC 100
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3360 AAGAACGCTTTGTCTGCTGCAGTGGAGGTGAGACGTTC 3398

RESULT 10
US-09-144-085-3
; Sequence 3, Application US/09144085
; Patent No. 6280999
; GENERAL INFORMATION:
; APPLICANT: Gustafsson, Claes
; APPLICANT: Betlach, Mary C.
; APPLICANT: Ashley, Gary
; APPLICANT: Julien, Bryan
; APPLICANT: Ziemann, Rainer
; TITLE OF INVENTION: SORANGIUM POLYKETIDE SYNTHASES AND ENCODING DNA
; TITLE OF INVENTION: THEREFOR
; FILE REFERENCE: 30062-20020.20
; CURRENT APPLICATION NUMBER: US/09/144,085
; EARLIER FILING DATE: 1998-08-31
; EARLIER APPLICATION NUMBER: 09/010,809
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 33529
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-144-085-3

Query Match          44.8%; Score 45.2; DB 4; Length 33529;
Best Local Similarity 66.3%; Pred. No. 0.00038;
Matches 65; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

OY 2 TCGACGCTGCTGGCGGCGGACCGCGACGGCTTCGACAGAGTCTCTGCGATCCTTGCGGCA 61
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1843 TCGACGCTGCTGGCGGCTCGGTGACGTGAGGCGGTGAGGAGCGGCTATGGGAGGCTGACGCA 1902

OY 62 AAGGGGGTGTCTGCGCTGCAGTCAAGGTGAGACGTCC 99
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1903 GAAGGTGTCTGCGCGGAGGTGAGACGTGACTACGC 1940

RESULT 11
US-09-103-840A-1/C
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00

```


;; CURRENT APPLICATION NUMBER: US/09/103, 840A
;; CURRENT FILING DATE: 1998-06-24
;; NUMBER OF SEQ. ID NOS: 2
;; SOFTWARE: PatentIn Ver. 2.1
;; SEQ ID NO 1
;; LENGTH: 4411529
;; TYPE: DNA
;; ORGANISM: Mycobacterium tuberculosis
;; OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match 35.2%; Score 35.6; DB 4; Length 4411529;
Best Local Similarity 60.2%; Pred. No. 0.31;
Matches 59; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

Oy 3 CGACGCTCTGGCGGCGACCGCGCTCGCAGAGTGTGCGCATCTTGGCGCA 62
Db 2304829 CGGTGGTGTCTCGGGAAGTGGCGCGCTGGAGAGCTGATCGTGTCCACCA 2304770

Oy 63 AGGGGCTTCTGCCGTGAGTCAAGTGAAGTCGCC 100
Db 2304769 AGGACTGCGACCCCGCGATCGAGTGAATATGCC 2304732

RESULT 12
US-09-428-517-1
; Sequence 1, Application US/09428517
; Patent No. 6251636
; GENERAL INFORMATION:
; APPLICANT: Bellach, Mary C.
; APPLICANT: Shai, Sanjay Krishnakant
; APPLICANT: McDaniel, Robert
; APPLICANT: Tang, Li
; TITLE OF INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 30062-20029.00
; CURRENT APPLICATION NUMBER: US/09/428, 517
; CURRENT FILING DATE: 1999-10-28
; EARLIER APPLICATION NUMBER: 60/120, 254
; EARLIER FILING DATE: 1998-02-16
; EARLIER APPLICATION NUMBER: 60/106, 100
; EARLIER FILING DATE: 1998-10-29
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 50937
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Recombinant DNA
US-09-428-517-1

Query Match 34.7%; Score 35; DB 4; Length 50937;
Best Local Similarity 59.6%; Pred. No. 0.26;
Matches 59; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

Oy 2 TCGACGGTCTGGGGGCGAGCGCGCTCGCAGAGTGTGGGCGATCTTGGCGCA 61
Db 20439 TCGACGGTGTGTCTGGGAGCGCCCGCTGGATGATGTGGCGGCTGTGAGCG 20498

Oy 62 AAGGGGTGTCTCTCGCTCGATCAAGGTGAGCTGCC 100
Db 20499 GAGGGGTCTCGGGTCTGTCTCGCTGTGACTATGCC 20537

RESULT 13
US-08-258-261B-6
; Sequence 6, Application US/08258261B
; Patent No. 5639949
; GENERAL INFORMATION:
; APPLICANT: Schnupp, Thomas
; APPLICANT: Ligon, James M.
; APPLICANT: Beck, James Joseph
; APPLICANT: Hill, Dwight Steven

;; APPLICANT: Ryals, John Andrew
;; APPLICANT: Gaffney, Thomas Deane
;; APPLICANT: Lam, Stephen Ting
;; APPLICANT: Hammer, Phillip E.
;; APPLICANT: Uknes, Scott Joseph
;; TITLE OF INVENTION: Genes for the synthesis of
;; TITLE OF INVENTION: antipathogenic substances
;; NUMBER OF SEQUENCES: 22
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Ciba-Geigy Corporation
;; STREET: 7 Skyline Drive
;; CITY: Hawthorne
;; STATE: NY
;; COUNTRY: USA
;; ZIP: 10532

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/258, 261B
FILING DATE: 08-JUN-1994
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/457, 205
FILING DATE: 01-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Elmer, James Scott
REGISTRATION NUMBER: 36,129
REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8614
TELEFAX: 919-541-8689

INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 28958 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-258-261B-6

Query Match 33.7%; Score 34; DB 1; Length 28958;
Best Local Similarity 59.2%; Pred. No. 0.46;
Matches 58; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

Oy 3 CGACGCTCTGGCGGCGACCGCGCTCGCAGAGTGTGCGCATCTTGGCGCA 62
Db 16640 CCACCTCTGATTCGCGCGACCGCGCGCTGAGAGCTGATGATCGCTCACCGGAG 16699
Oy 63 AGGGGCTTCTGCCGTGAGTCAAGTGAAGTCGCC 100
Db 16700 CGACGCTTCTGCCCGCGAGAGTCCCGCTGACTATGCC 16737

RESULT 14
US-08-456-837-6
; Sequence 6, Application US/08456837
; Patent No. 5643774
; GENERAL INFORMATION:
; APPLICANT: Schnupp, Thomas
; APPLICANT: Ligon, James M.
; APPLICANT: Beck, James Joseph
; APPLICANT: Hill, Dwight Steven
; APPLICANT: Ryals, John Andrew
; APPLICANT: Gaffney, Thomas Deane
; APPLICANT: Lam, Stephen Ting
; APPLICANT: Hammer, Phillip E.
; APPLICANT: Uknes, Scott Joseph
; TITLE OF INVENTION: Genes for the synthesis of

```

; TITLE OF INVENTION: antipathogenic substances
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/456,837
; FILING DATE: 01-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/457,205
; FILING DATE: 01-JUN-1995
; APPLICATION NUMBER: 08/258,261
; FILING DATE: 08-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8614
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28958 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-456-837-6

Query Match          33.7%; Score 34; DB 1; Length 28958;
Best Local Similarity 59.2%; Pred. No. 0.46;
Matches 58; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 3 CGACGCTGTCGGCGGCGGAGCGCGCTGCGAGAGTCGCGGAGTCCTTGGCGCAA 62
DB 16640 CCACGCTGCTATCCGCGGAGCCCGCGCGCTGCGAGTCGACTGCTCACCAG 16699
QY 63 AGGGGCTGTTCTGCGTCGAGTCAGAGTGAGTGCCTGCC 100
DB 16700 CGAGGCTTTCGCCGCGAAGAGTCGCGCTGACTACGCC 16737

RESULT 15
US-08-457-342-6
; Sequence 6, Application US/08457342
; Patent No. 5662898
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James M.
; APPLICANT: Beck, James Joseph
; APPLICANT: Hill, Dwight Steven
; APPLICANT: Ryals, John Andrew
; APPLICANT: Gaffney, Thomas Deane
; APPLICANT: Lam, Stephen Ting
; APPLICANT: Hammer, Phillip E.
; APPLICANT: Uknes, Scott Joseph
; TITLE OF INVENTION: Genes for the synthesis of
; TITLE OF INVENTION: antipathogenic substances
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
```

```

; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/457,342
; FILING DATE: 01-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/457,205
; FILING DATE: 01-JUN-1995
; APPLICATION NUMBER: 08/258,261
; FILING DATE: 08-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8614
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28958 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-457-342-6

Query Match          33.7%; Score 34; DB 1; Length 28958;
Best Local Similarity 59.2%; Pred. No. 0.46;
Matches 58; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 3 CGACGCTGTCGGCGGCGGAGCGCGCTGCGAGAGTCGCTGCGGAGTCCTTGGCGCAA 62
DB 16640 CCACGCTGCTATCCGCGGAGCCCGCGCGCTGCGAGTCGACTGCTCACCAG 16699
QY 63 AGGGGCTGTTCTGCGTCGAGTCAGAGTGAGTGCCTGCC 100
DB 16700 CGAGGCTTTCGCCGCGAAGAGTCGCGCTGACTACGCC 16737
```

Search completed: November 6, 2002, 16:26:00
Job time : 1248.62 secs


```
FT FT /tag= d /label= epoc_gene
FT FT /note= "encodes ketide synthase (KS-Y) of the loading
FT FT domain"
FT FT 3621..4661 /note= "encodes module 2"
FT FT /tag= y
FT FT /note= "encodes KS2, the KS domain of module 2"
FT FT 12250..13287
FT FT /tag= z
FT FT /note= "encodes AT2, the AT domain of module 2"
FT FT 13327..13899
FT FT /tag= aa
FT FT /note= "encodes dehydratase (DH) 2, the DH domain of
FT FT module 2"
FT FT 14962..15756
FT FT /tag= ab
FT FT /note= "encodes ketoreductase (KR) 2, the KR domain of
FT FT module 2"
FT FT 15763..16008
FT FT /tag= ac
FT FT /note= "encodes ACP2, the ACP domain of module 2"
FT FT 16134..37907
FT FT /tag= ad
FT FT /label= epod_gene
FT FT /note= "encodes modules 3-6"
FT FT 16425..17606
FT FT /tag= ae
FT FT /note= "encodes KS3"
FT FT 17817..18857
FT FT /tag= af
FT FT /note= "encodes AT3"
FT FT 19581..20396
FT FT /tag= ag
FT FT /note= "encodes KR3"
FT FT 20424..20642
FT FT /tag= ah
FT FT /note= "encodes ACP3"
FT FT 20706..22082
FT FT /tag= ai
FT FT /note= "encodes KS4"
FT FT 22296..23336
FT FT /tag= aj
FT FT /note= "encodes AT4"
FT FT 24069..24647
FT FT /tag= ak
FT FT /note= "encodes KR4"
FT FT 24867..25151
FT FT /tag= al
FT FT /note= "encodes ACP4"
FT FT 25203..26576
FT FT /tag= am
FT FT /note= "encodes KS5"
FT FT 26793..27883
FT FT /tag= an
FT FT /note= "encodes AT5"
FT FT 27966..28574
FT FT /tag= ao
FT FT /note= "encodes DH5"
FT FT 29433..30287
FT FT /tag= ap
FT FT /note= "encodes ER5"
FT FT 30321..30869
FT FT /tag= aq
FT FT /note= "encodes KR5"
FT FT 31077..31373
FT FT /tag= ar
FT FT /note= "encodes ACP5"
FT FT 31440..32807
FT FT /tag= as
FT FT /note= "encodes KS6"
FT FT 33018..34067
FT FT /tag= at
FT FT /note= "encodes AT6"
FT FT 34107..34676
FT FT /tag= au

FT FT /tag= d
FT FT /note= "encodes ketide synthase (KS-Y) of the loading
FT FT domain"
FT FT 3621..4661
FT FT /tag= e
FT FT /note= "encodes acyl transferase (AT) of the loading
FT FT domain"
FT FT 4917..5810
FT FT /tag= f
FT FT /note= "encodes enoyl reductase (ER) of the loading
FT FT domain, potentially involved in formation of the
FT FT thiazole moiety"
FT FT 5856..6155
FT FT /tag= g
FT FT /note= "encodes acyl carrier protein (ACP) of the loading
FT FT domain"
FT FT 6260..10493
FT FT /tag= h
FT FT /label= epod_gene
FT FT /note= "encodes module 1, the NRPS module"
FT FT 2031..3548
FT FT /tag= i
FT FT /note= "encodes condensation domain C2 of the NRPS
FT FT module"
FT FT 2031..3548
FT FT /tag= j
FT FT /note= "encodes condensation domain C2 of the NRPS
FT FT module"
FT FT 2031..3548
FT FT /tag= k
FT FT /note= "encodes heterocyclization signature sequence"
FT FT 6861..6887
FT FT /tag= l
FT FT /note= "encodes condensation domain C4 of the NRPS
FT FT module"
FT FT 7358..7366
FT FT /tag= m
FT FT /note= "encodes condensation domain C7 (partial) of the
FT FT NRPS module"
FT FT 7898..7921
FT FT /tag= n
FT FT /note= "encodes adenylation domain A1 of the NRPS module"
FT FT 7898..7921
FT FT /tag= o
FT FT /note= "encodes adenylation domain A1 of the NRPS module"
FT FT 8261..8308
FT FT /tag= p
FT FT /note= "encodes adenylation domain A3 of the NRPS module"
FT FT 8411..8422
FT FT /tag= q
FT FT /note= "encodes adenylation domain A4 of the NRPS module"
FT FT 8861..8905
FT FT /tag= r
FT FT /note= "encodes adenylation domain A6 of the NRPS module"
FT FT 8966..8983
FT FT /tag= s
FT FT /note= "encodes adenylation domain A7 of the NRPS module"
FT FT 9090..9179
FT FT /tag= t
FT FT /note= "encodes adenylation domain A8 of the NRPS module"
FT FT 9183..9992
FT FT /tag= u
FT FT /note= "encodes oxidation region for forming thiazole"
FT FT 10121..10138
FT FT /tag= v
FT FT /note= "encodes adenylation domain A10 of the NRPS
FT FT module"
FT FT 10261..10306
FT FT /tag= w
FT FT /note= "encodes thiolation domain (PCP) of the NRPS
FT FT module"
FT FT 10639..16137
FT FT /tag= x
```


XX	WP1: 2000-097741.08	
DR	P-PSDB: AAY58573, AAY58574, AAY58575, AAY58576, AAY58577, AAY58578,	
DR	AAV58579, AAY58580, AAY58581, AAY58582, AAY58583, AAY58584,	
DR	AAV58585, AAY58586, AAY58587, AAY58588, AAY58590, AAY58591,	
XX	AAV58592, AAY58593, AAY58594.	
PT	New isolated epothilone synthase genes, used for the recombinant	
PT	production of epothilone for use in cancer therapy	
XX	Claim 14; Page 87-104; 174pp; English.	
XX	This sequence represents a 68.75 kb contig from Sorangium cellulosum	
CC	comprising 22 open reading frames (ORFs) and includes genes encoding	
CC	proteins involved in the biosynthesis of epothilones. Epothilones A and	
CC	B are 16-membered macrocyclic polyketides with an acylcysteine-derived	
CC	starter unit; polyketides being synthesised from two-carbon building	
CC	blocks, the beta-carbon of which always carries a keto group. Each round	
CC	of two-carbon addition is carried out by a complex of enzymes known as	
CC	the polyketide synthase in a manner similar to fatty acid biosynthesis.	
CC	EPoS A (AAV58573) and EPoS P (AAV58574) are involved in formation of	
CC	the thiazole ring formation of epothilones, and EPoS B, EPoS C, EPoS D	
CC	and EPoS E (AAV58575-Y58578) are involved in polyketide backbone	
CC	formation. EPo F (AAV58579) is an epothilone macrolactone oxidase, and	
CC	the proteins Orf 3 (AAV58582) and Orf14 (AAV58593) are thought to be	
CC	involved in transport. Epothilones mimic the biological activity of	
CC	taxol, and may be substituted for taxol in cancer chemotherapeutic	
CC	compositions. Epothilones exhibit a much lower drop in potency against a	
CC	multiply drug-resistant cell line compared with taxol, and are	
CC	considerably less efficiently exported from such cells by the multidrug	
CC	resistance protein (MDR, or P-glycoprotein). Despite the potential of	
CC	epothilones as anticancer agents, they are problematical to produce on a	
CC	large scale. Epothilones are too complex for industrial scale chemical	
CC	synthesis, and Sorangium cellulosum is difficult to ferment, producing	
CC	poor yields of epothilones. The nucleic acids of the invention may be	
CC	used for the recombinant production of epothilones in a heterologous host	
CC	that is more amenable to fermentation.	
XX	Sequence 68750 BP; 9596 A; 22458 C; 25337 G; 11159 T; 0 other;	
SO		
Query Match	99.0%; Score 100; DB 21; Length 68750;	
Best Local Similarity	100.0%; Pred. No. 2,7e-18;	
Matches	100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
OY	2 TCGACGGTCTGCGCGGAGCCGAGCCGATCGCAGAGAGTCTGGCGATCCTTGGCGCA 61	
Db	39113 TCGACGGTCTGCGCGGAGCCGAGCCGATCGCAGAGAGTCTGGCGATCCTTGGCGCA 39172	
OY	62 AAGGGGGTCTTCTGCCGTCGACTCAAGGTGACGTCGCCCA 101	
Db	39173 AAGGGGGTCTTCTGCCGTCGACTCAAGGTGACGTCGCCCA 39212	
RESULT 3		
AAV05287		
ID	AAV05287 standard; DNA; 49377 BP.	
XX		
XX	AAV05287;	
XX		
DT	21-MAY-1998 (first entry)	
DE	The soraphen biosynthesis gene cluster from Sorangium cellulosum.	
XX		
KW	Polyketide synthase; PKS; biosynthesis; soraphen; SorP; SorA; SorB;	
KW	SorM; biosynthetic module; beta-ketoacyl synthase; acyltransferase;	
KW	ketoreductase; beta-ketone processing domain; cytosolic agent;	
KW	antimicrobial agent; phytopathogenic fungi; transgenic plant;	
XX	biological control; ss.	
CS		
XX	Sorangium cellulosum.	
XX		
Key	Location/Qualifiers	
TH		
CD	383..760	
CD		

FT		/*tag= a	
FT		/product= SorB	
FT		/note= "gene product highly homologous to the	
FT		reductase domains of type I PKsS such	
FT		as eryA from Saccharopolyspora erythraea"	
FT			
FT	CDS	927..19874	
FT		/*tag= b	
FT		/product= SorA	
FT		/note= "gene product is highly homologous to	
FT		type I PKsS that are known to be involved	
FT		in the synthesis of polyketide compounds"	
FT		942..7115	
FT		/*tag= c	
FT		/note= "module 1 of SorA"	
FT		7203..12884	
FT		/*tag= d	
FT		/note= "module 2 of SorA"	
FT		13455..19616	
FT		/*tag= e	
FT		/note= "module 3 of SorA"	
FT		19871..46318	
FT		/*tag= f	
FT		19871..46318	
FT	CDS		
FT		/*tag= g	
FT		/product= SorB	
FT		/note= "gene product is highly homologous to	
FT		type I PKs genes"	
FT		19870..24556	
FT		/*tag= h	
FT		/note= "module 1 of SorB"	
FT		24638..30820	
FT		/*tag= i	
FT		/note= "module 2 of SorB"	
FT		30881..35446	
FT		/*tag= j	
FT		/note= "module 3 of SorB"	
FT		35528..40114	
FT		/*tag= k	
FT		/note= "module 4 of SorB"	
FT		40190..46318	
FT		/*tag= l	
FT		/note= "module 5 of SorB"	
FT	CDS	46851..47891	
FT		/*tag= m	
FT		/product= SorM	
FT		/note= "gene product is homologous to the	
FT		methyltransferase from Streptomyces	
FT		hygroscopicus that is involved in	
FT		the synthesis of the polyketide rappaamc1a"	
XX			
XX			
XX			
XX	US5716849-A.		
XX			
XX	10-FEB-1998.		
XX			
XX	14-DEC-1996;	96US-0764233.	
XX			
XX	24-AUG-1993;	93WO-US07954.	
XX	08-JUN-1994;	94US-0258261.	
XX	09-OCT-1996;	96US-0729214.	
XX			
XX	(NOVS) NOVARTIS FINANCE CORP.		
XX			
XX	Beck JY, Hill DS, Ligon JM, Neff S, Ryals JA, Schupp T;		
XX			
XX	WPI: 1998-158369/14.		
XX			
XX	DNA encoding Sorangium cellulosum polypeptide(s) - used for, e.g.		
XX	biosynthesis of soraphen useful as antimicrobial agent against		
XX	phytopathogenic fungi		
XX			
XX	Claim 2; Columns 47-90; 64pp; English.		
XX			
XX			
XX	The present sequence contains a cluster of genes that encode polyketide		

CC synthases (PKSs) that are involved in the synthesis of soraphens in
CC Sorangium cellulosum. The proteins encoded by the present sequence are
CC SorA, SorB, SorC and SorM. SorA and SorB contain biosynthetic modules
CC which contain a beta-ketoacyl synthase, an acyltransferase, a
CC ketoreductase and an acyl carrier protein domain, as well as beta-ketone
CC processing domains. S. cellulosum soraphens are useful as a cytosstatic
CC and antimicrobial agent active against phytopathogenic fungi.
CC Soraphen-producing transgenic plants or biological control agents can
CC also be produced, which may reduce crop losses and nutritional
CC deprivation for local populations in many parts of the world.
XX
SQ Sequence 49377 BP; 7247 A; 19522 C; 14477 G; 8131 T; 0 other;
Query Match 52.1%; Score 52.6; DB 19; Length 49377;
Best Local Similarity 70.7%; Pred. No. 2.6e-05;
Matches 70; Conservative 0; Mismatches 29; Indels 0; Gaps 0;
QY 2 TCAGAGGTGCTGCGGGGCGAGCCGCGCTCCAGAGTGTCTGCGATCTTGGCGCA 61
Db 3300 TCACGGGTGCTGGGGGGGAGCGGAGCGGCTCGAGCGGCTGTCAGAGCACTGAGCGG 3359
QY 62 AAGGGGTGTTCTGCGCTCGAGTCAAGGTGAGCTGCC 100
Db 3360 AAGACGCTTCTTGTGCTGCGAGTGGCGATGACGTTGCC 3398
RESULT 4
AAF90036
ID AAF90036 standard; DNA; 5292 BP.
XX
AC AAF90036;
XX
DT 06-AUG-2001 (first entry)
DE Nucleotide sequence of a type I polyketide synthase.
XX
KM Metabolic pathway operon; polyketide; polyketide antibiotic;
KM type I polyketide synthase; ss.
XX
OS Unidentified.
XX
FH Key Location/Qualifiers
FT CDS 1..5292
FT /tag= a
FT /product= "type I polyketide synthase"
XX
PN WO200140497-A2.
XX
PD 07-JUN-2001.
XX
PF 27-NOV-2000; 2000MO-FR03311.
XX
PR 29-NOV-1999; 99FR-0015032.
PR 07-JUN-2000; 2000US-0209800.
XX
PA (AVENTIS PHARMA SA.
XX
PI Jeanin P, Pernodet J, Guerinneau M, Simonet P, Courtois S;
PI Cappellano C, Francou F, Raynal A, Ball M, Sezonov G, Tuppille K;
PI Frostegard A;
XX
DR WPI: 2001-374849/39.
DR P-PDB: AAB83973.
XX
XX Collection of nucleic acids from environmental samples, useful for
PT identifying e.g. genes encoding polyketide synthases and derived
PT antibiotics
XX
PS Claim 35; Page 305-307; 356pp; French.
XX
CC The specification describes a method for the preparation of a collection
CC of nucleic acids from organisms in a soil sample. The method comprises
CC milling a dried sample to produce microparticles; suspending these in

CC liquid buffer; extraction of nucleic acids from the microparticle;
CC passing nucleic acid-containing solution through a molecular sieve;
CC passing nucleic acid-enriched fractions through an anion exchange
CC chromatography material; and recovering fractions containing purified
CC nucleic acids. The nucleic acids are sources for sequences that encode
CC either operons involved in a metabolic pathway (specifically polyketide
CC synthasis) or polypeptides, particularly for production of therapeutic
CC or agricultural compounds, especially polyketide antibiotics. AAF90034-39
CC represent open reading frames (ORFs) of the coding strand of cosmid
CC a26g1, and encode type I polyketide synthases.
XX
SQ Sequence 5292 BP; 905 A; 1715 C; 1725 G; 947 T; 0 other;
Query Match 47.5%; Score 48; DB 22; Length 5292;
Best Local Similarity 68.8%; Pred. No. 0.00038;
Matches 66; Conservative 0; Mismatches 30; Indels 0; Gaps 0;
QY 5 ACGGTGCTGCGGGGCGAGCCGCGCTCCAGAGTGTCTGCGATCTTGGCGCAAG 64
Db 2134 ACCGTCCTGTGGGCGAGCGAGCACTGTGGGCGCAACTGCTGGGGAACCTGAGCGCAA 2193
QY 65 GGGGTGTTCTGCGCTCGAGTCAAGGTGAGCTGCC 100
Db 2194 GACGCTTCTGCGCTCGCGGTGAAGTGACATTTGCC 2229
RESULT 5
AAF90033
ID AAF90033 standard; DNA; 34071 BP.
XX
AC AAF90033;
XX
DT 06-AUG-2001 (first entry)
DE Nucleotide sequence of cosmid a26g1 (coding strand).
XX
KM Metabolic pathway operon; polyketide; polyketide antibiotic; ss.
KM Synthetic.
XX
OS Synthetic.
XX
PN WO200140497-A2.
XX
PD 07-JUN-2001.
XX
PF 27-NOV-2000; 2000MO-FR03311.
XX
PR 29-NOV-1999; 99FR-0015032.
PR 07-JUN-2000; 2000US-0209800.
XX
PA (AVENTIS PHARMA SA.
XX
PI Jeanin P, Pernodet J, Guerinneau M, Simonet P, Courtois S;
PI Cappellano C, Francou F, Raynal A, Ball M, Sezonov G, Tuppille K;
PI Frostegard A;
XX
DR WPI: 2001-374849/39.
XX
XX Collection of nucleic acids from environmental samples, useful for
PT identifying e.g. genes encoding polyketide synthases and derived
PT antibiotics
XX
PS Example 14; Page 289-300; 356pp; French.
XX
XX The specification describes a method for the preparation of a collection
CC of nucleic acids from organisms in a soil sample. The method comprises
CC milling a dried sample to produce microparticles; suspending these in
CC liquid buffer; extraction of nucleic acids from the microparticle;
CC passing nucleic acid-containing solution through a molecular sieve;
CC passing nucleic acid-enriched fractions through an anion exchange
CC chromatography material; and recovering fractions containing purified
CC nucleic acids. The nucleic acids are sources for sequences that encode
CC either operons involved in a metabolic pathway (specifically polyketide
CC synthasis) or polypeptides, particularly for production of therapeutic

CC or agricultural compounds, especially polyketide antibiotics. The present
CC sequence represents cosmid a26g1 coding strand), which encodes different
CC polyketide synthases.

SO Sequence 34071 BP: 5791 A; 10858 C; 11089 G; 6333 T; 0 other;

Query Match

Best Local Similarity 47.5%; Score 48; DB 22; Length 34071;

Best Local Similarity 68.8%; Pred. No. 0.00046;

Matches 66; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

OY 5 ACGGTGCTGGCGGCGAGCGCGTCGCGAGAGTGTGTCGCAATCTTGCGGCAAG 64

DB 15069 ACCGTCTGTGCGGCGACACCGACGCTGTGGCGAACTGCTGCGGAACCTGAGCGGAAA 15128

OY 65 GGGGTGTCTGCCGTCGAGTCAGTGTGACGTGCC 100

DB 15129 GACGTCTTCTGCGCGTCGCGTGAAGTGACATTTGCC 15164

RESULT 6

AAF90032/C
ID AAF90032 standard; DNA: 42717 BP.

AC AAF90032;

DT 06-AUG-2001 (first entry)

DE Nucleotide sequence of cosmid a26g1 (non-coding strand).

KW Metabolic pathway operon: polyketide; polyketide antibiotic; ss.

OS Synthetic.

PN WO200140497-A2.

PD 07-JUN-2001.

PF 27-NOV-2000; 2000WO-FR0311.

PR 29-NOV-1999; 99FR-0015032.

PR 07-JUN-2000; 2000US-0209800.

PA (AVET) AVENTIS PHARMA SA.

PI Jeanmin P, Pernodet J, Guerinneau M, Simonet P, Courtois S;

PI Cappellano C, Francou F, Raynal A, Ball M, Sezonov G, Tiphille K;

PI Frosteberg A;

DR WPI; 2001-374849/39.

XX Collection of nucleic acids from environmental samples, useful for

PT identifying e.g. genes encoding polyketide synthases and derived

PT antibiotics

XX Example 14; Page 274-288; 356pp; French.

CC The specification describes a method for the preparation of a collection

CC of nucleic acids from organisms in a soil sample. The method comprises

CC milling a dried sample to produce microparticles; suspending these in

CC a liquid buffer; extraction of nucleic acids from the microparticle;

CC passing nucleic acid-containing solution through a molecular sieve;

CC chromatography material; and recovering fractions containing purified

CC nucleic acids. The nucleic acids are sources for sequences that encode

CC either operons involved in a metabolic pathway (specifically polyketide

CC synthesis) or polypeptides, particularly for production of therapeutic

CC or agricultural compounds, especially polyketide antibiotics. The present

CC sequence represents cosmid a26g1 (non-coding strand). The sense strand

CC encodes different polyketide synthases.

XX Sequence 42717 BP: 8230 A; 13520 C; 13184 G; 7782 T; 1 other;

SO Query Match 47.5%; Score 48; DB 22; Length 42717;

Best Local Similarity 68.8%; Pred. No. 0.00047;

Matches 66; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

OY 5 ACGGTGCTGGCGGCGAGCGCGTCGCGAGAGTGTGTCGCAATCTTGCGGCAAG 64

DB 19051 ACCGTCTGTGCGGCGACACCGACGCTGTGGCGAACTGCTGCGGAACCTGAGCGGAAA 18992

OY 65 GGGGTGTCTGCCGTCGAGTCAGTGTGACGTGCC 100

DB 18991 GACGTCTTCTGCGCGTCGCGTGAAGTGACATTTGCC 18956

RESULT 7

AAAI4651/C
ID AAI4651 standard; DNA: 77536 BP.

AC AAI4651;

DT 08-AUG-2000 (first entry)

DE Nucleotide sequence of the FK-520 biosynthetic gene cluster.

KW FK-520; polyketide synthase; PKS; gene cluster; immunosuppressant;

KW Streptomyces hygroscopicus var. ascomyceticus; immunophilin;

KW FK-506 binding protein; polyketide compound; transplant rejection;

KW graft-versus-host disease; uveitis; alopecia universalis;

KW autoimmune chronic active hepatitis; inflammatory bowel disease;

KW multiple sclerosis; primary biliary cirrhosis; scleroderma;

KW neurite outgrowth; nerve regrowth; Parkinson's disease;

KW Alzheimer's disease; stroke; traumatic spinal cord; brain injury;

KW peripheral neuropathy; ss.

OS Streptomyces hygroscopicus.

PN location/Qualifiers

FT complement (412..1836)

FT /tag= a

FT /note= "fkbp gene"

FT complement (2020..3579)

FT /tag= b

FT /note= "fkbp gene"

FT complement (3969..4496)

FT /tag= c

FT /note= "fkbr2 gene"

FT complement (4595..5488)

FT /tag= d

FT /note= "fkbr1 gene"

FT complement (5601..6818)

FT /tag= e

FT /note= "fkbr gene"

FT complement (6808..8052)

FT /tag= f

FT /note= "fkbr gene"

FT complement (8156..8824)

FT /tag= g

FT /note= "fkbg gene"

FT complement (9122..9883)

FT /tag= h

FT /note= "fkbg gene"

FT complement (9894..10994)

FT /tag= i

FT /note= "fkbi gene"

FT complement (10987..11247)

FT /tag= j

FT /note= "fkbi gene"

FT complement (11244..12092)

FT /tag= k

FT /note= "fkpk gene"

FT complement (12113..13150)

FT /tag= l

FT /note= "fkbl gene"

FT complement (13212..23988)

FT /tag= m

FT /note= "fkbc gene"
FT complement (13452..13662)
FT /*tag= n
FT /note= "ACP6"
FT complement (13761..14394)
FT /*tag= o
FT /note= "KR6"
FT complement (14517..15294)
FT /*tag= p
FT /note= "ER6"
FT complement (15438..16587)
FT /*tag= q
FT /note= "dehydratase domain (DH) 6"
FT complement (16587..17820)
FT /*tag= r
FT /note= "acyltransferase domain (AT) 6"
FT complement (17820..19053)
FT /*tag= s
FT /note= "KS6"
FT complement (19116..19326)
FT /*tag= t
FT /note= "ACP5"
FT complement (19464..20097)
FT /*tag= u
FT /note= "KR5"
FT complement (20241..21420)
FT /*tag= v
FT /note= "DH5"
FT complement (21420..22653)
FT /*tag= w
FT /note= "AT5"
FT complement (22653..23892)
FT /*tag= x
FT /note= "KS5"
FT complement (23992..46573)
FT /*tag= y
FT /note= "fkbc gene"
FT complement (24163..24373)
FT /*tag= z
FT /note= "ACP4"
FT complement (24997..26146)
FT /*tag= aa
FT /note= "DH4 (inactive)"
FT complement (26146..27430)
FT /*tag= ab
FT /note= "AT4"
FT complement (27430..28684)
FT /*tag= ac
FT /note= "KS4"
FT complement (28750..28960)
FT /*tag= ad
FT /note= "ACP3"
FT complement (29092..29740)
FT /*tag= ae
FT /note= "KR3"
FT complement (29869..31018)
FT /*tag= af
FT /note= "DH3 (inactive)"
FT complement (31018..32185)
FT /*tag= ag
FT /note= "AT3"
FT complement (32185..33439)
FT /*tag= ah
FT /note= "KS3"
FT complement (33505..33715)
FT /*tag= ai
FT /note= "ACP2"
FT complement (33823..34480)
FT /*tag= aj
FT /note= "KR2"
FT complement (34606..35749)
FT /*tag= ak
FT complement (28750..28960)
FT misc_feature

FT /*tag= al
FT /note= "DH2 (inactive)"
FT complement (35749..37144)
FT /*tag= am
FT /note= "AT2"
FT complement (37145..38296)
FT /*tag= an
FT /note= "KS2"
FT complement (38371..38581)
FT /*tag= ao
FT /note= "ACP1"
FT complement (38677..39307)
FT /*tag= ap
FT /note= "KR1"
FT complement (39442..40609)
FT /*tag= aq
FT /note= "DH1"
FT complement (40609..41842)
FT /*tag= ar
FT /note= "AT1"
FT complement (41842..43093)
FT /*tag= as
FT /note= "KS of extender module 1 (KS1)"
FT complement (43144..43660)
FT /*tag= at
FT /note= "ACP of loading domain"
FT complement (43777..44629)
FT /*tag= au
FT /note= "ER of loading domain"
FT complement (44974..46573)
FT /*tag= av
FT /note= "CoA ligase of loading domain"
FT 46754..47788
FT /*tag= aw
FT /note= "fkbc gene"
FT 47785..52272
FT /*tag= ax
FT /note= "fkbp gene"
FT 52275..71465
FT /*tag= ay
FT /note= "fkba gene"
FT 52362..53576
FT /*tag= az
FT /note= "KS7"
FT 53577..54716
FT /*tag= bb
FT /note= "AT7"
FT 54717..55871
FT /*tag= ba
FT /note= "DH7"
FT 56019..56819
FT /*tag= bc
FT /note= "ER7"
FT 56943..57575
FT /*tag= bd
FT /note= "KR7"
FT 57710..57920
FT /*tag= be
FT /note= "ACP7"
FT 57990..59243
FT /*tag= bf
FT /note= "KS6"
FT 59244..60398
FT /*tag= bg
FT /note= "AT6"
FT 60399..61412
FT /*tag= bh
FT /note= "DH8 (inactive)"
FT misc_feature

Query Match 47.3%; Score 47.8; DB 21; Length 77536;
Best Local Similarity 67.7%; Pred. No. 0.00056;
Matches 67; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

FT / *tag = m

```
FT      misc_feature      complement (28750..28960)
```

```

FT      /tag= a1
FT      /note= "DH2 (inactive)"
FT      /complement (35749..37144)
FT      /tag= am
FT      /note= "AT2"
FT      /complement (37145..38296)
FT      /tag= an
FT      /note= "KS2"
FT      /complement (38371..38581)
FT      /tag= ao
FT      /note= "ACPI"
FT      /complement (38677..39207)
FT      /tag= ap
FT      /note= "KR1"
FT      /complement (39442..40609)
FT      /tag= aq
FT      /note= "DH1"
FT      /complement (40609..41842)
FT      /tag= ar
FT      /note= "AT1"
FT      /complement (41842..43093)
FT      /tag= as
FT      /note= "KS of extender module 1 (KSI)"
FT      /complement (43144..43660)
FT      /tag= at
FT      /note= "ACP of loading domain"
FT      /complement (43777..44629)
FT      /tag= au
FT      /note= "ER of loading domain"
FT      /complement (44974..46573)
FT      /tag= av
FT      /note= "COA ligase of loading domain"
FT      /tag= aw
FT      /note= "fkbD gene"
FT      /tag= ax
FT      /note= "fkbP gene"
FT      /tag= ay
FT      /note= "fkbA gene"
FT      /tag= az
FT      /note= "KS7"
FT      /tag= ba
FT      /note= "AT7"
FT      /tag= bc
FT      /note= "ER7"
FT      /tag= bd
FT      /note= "KR7"
FT      /tag= be
FT      /note= "ACPI"
FT      /tag= bf
FT      /note= "KS8"
FT      /tag= bg
FT      /note= "AT8"
FT      /tag= bh
FT      /note= "DH (inactive)"

```

Query Match 37.8%; Score 38.2; DB 21; Length 77536;
 Best Local Similarity 61.6%; Pred. No. 0.24; Indels 0; Gaps 0;
 Matches 61; Conservative 0; Mismatches 38;

```

QY      2 TCGACGGTCTGCGGCGAGCCGCGAGCGCTGCAGAGGTGCTGGCATTCTTGGCGCA 61
        ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      64557 TCGACAGTCTGTGCGCGCGCGGTGCGGTGAGAGAGTGTGACGCGGTATGAGACC 64616
QY      62 AAGGGGTGTCTGCGGTGAGTCAAGGTGAGTGGCC 100
        || |||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      64617 GAAGGCGTGGAGTGTGTATGCGCGGTGACTACGCC 64655

RESULT 11
AA14664
ID      AA14664 standard; DNA: 4547 BP.
XX
AC      AA14664;
XX
DT      08-AUG-2000 (first entry)
XX
DE      Nucleotide sequence of a fragment of the FK-520 PKS gene cluster.
XX
KW      FK-520; polyketide synthase; PKS; gene cluster; immunosuppressant;
KW      immunophilin; FK-506 binding protein; polyketide compound; uveitis;
KW      transplant rejection; graft-versus-host disease; alopecia universalis;
KW      autoimmune chronic active hepatitis; inflammatory bowel disease;
KW      multiple sclerosis; primary biliary cirrhosis; scleroderma;
KW      multiple growth; nerve regrowth; Parkinson's disease;
KW      Alzheimer's disease; stroke; traumatic spinal cord; brain injury;
KW      peripheral neuropathy; ss.
XX
OS      Synthetic.
OS      Streptomyces hygroscopicus.
XX
FH      Key Location/Qualifiers
FT      CDS 9..4535
FT      /tag= a
FT      /note= "no termination codon given"
XX
PN      WO200020601-A2.
XX
PD      13-APR-2000.
XX
PF      01-OCT-1999; 99WO-US22886.
XX
PR      02-OCT-1998; 98US-0102748.
PR      11-MAR-1999; 99US-0123810.
PR      17-JUN-1999; 99US-0139650.
XX
PA      (KOSA-) KOSAN BIOSCIENCES INC.
XX
PI      Reeves C, Chu D, Khosla C, Santl D, Wu K;
XX
DR      WPI; 2000-317716/27.
DR      P-PSDB; AAY84728.
XX
PT      New isolated polyketide synthase nucleic acid and polyketide compounds,
PT      useful for treating e.g. transplant rejection, uveitis, multiple
PT      sclerosis, Alzheimer's disease, Parkinson's disease, stroke, or
PT      peripheral neuropathy -
XX
PS      Example 1; Page 85-88; 126pp; English.
XX
CC      The present sequence represents an NheI-XhoII fragment that encodes
CC      module 8 of the FK-520 polyketide synthase (PKS) gene cluster with the
CC      endogenous acyltransferase (AT) domain replaced by the AT domain of
CC      module 13 of the rapamycin PKS (which is specific for methylmalonyl CoA).
CC      FK-506 is a potent immunosuppressant, and acts through initial formation
CC      of an intermediate complex with protein immunophilins known as FK-506
CC      binding proteins. The nucleic acids are used for producing polyketide
CC      compounds. The polyketide compounds can be used as immunosuppressants to
CC      prevent or treat transplant rejection, graft-versus-host disease or
CC      uveitis. They can also be used for treating e.g. alopecia universalis,
CC      autoimmune chronic active hepatitis, inflammatory bowel disease,
CC      multiple sclerosis, primary biliary cirrhosis, or scleroderma. They
CC      also have neurotrophic activity and can be used to promote neurite

```



```

XX (KOSA-) KOSAN BIOSCIENCES INC.
PA
XX
XX Reeves C, Chu D, Khosla C, Santl D, Wu K;
PI
XX WPI: 2000-317716/27.
XX P-PSDB: AAY84731.
DR
XX New isolated polyketide synthase nucleic acid and polyketide compounds,
PT useful for treating e.g. transplant rejection, uveitis, multiple
PT sclerosis, Alzheimer's disease, Parkinson's disease, stroke, or
PT peripheral neuropathy -
XX
XX Example 2; Page 96-99; 126pp: English.
PS
XX The present sequence represents module 8 of the FK-520 polyketide
CC synthase (PKS) gene cluster, containing the acyltransferase (AT)
CC domain of module 13 of rapamycin. FK-506 is a potent immunosuppressant,
CC and acts through initial formation of an intermediate complex with
CC protein immunophilins known as FK-506 binding proteins. The nucleic
CC acids are used for producing polyketide compounds. The polyketide
CC compounds can be used as immunosuppressants to prevent or treat
CC transplant rejection, graft-versus-host disease or uveitis. They can
CC also be used for treating e.g. alopecia universalis, autoimmune
CC chronic active hepatitis, inflammatory bowel disease, multiple
CC sclerosis, primary biliary cirrhosis, or scleroderma. They
CC also have neurotrophic activity and can be used to promote neurite
CC outgrowth in NGF-treated PC12 cells and in sensory neuronal cultures,
CC and in intact animals, they promote regrowth of damaged facial and
CC sciatic nerves, and repair lesioned serotonin and dopamine neurons in
CC the brain. They can also be used for treating e.g. Parkinson's disease,
CC Alzheimer's disease, stroke, traumatic spinal cord and brain injury, or
CC peripheral neuropathies. They can also be used in agricultural and
CC veterinary applications.
XX
XX Sequence 4767 BP; 731 A; 1945 C; 1468 G; 623 T; 0 other;
SO
Query Match 36.2%; Score 36.6; DB 21; Length 4767;
Best Local Similarity 60.6%; Pred. No. 0.5;
Matches 60; Conservative 0; Mismatches 39; Indels 0; Gaps 0;
OY 2 TCGACGGTCTGCGGGCGGCGAGCGCTGCGAGAGTCTGCGATCTTGCGGCA 61
DB 2454 TCCACCGTATGCGGGCGACCCCGGAAGCGGTGACCATGCTCTACCGCTATGAGCA 2513
OY 62 AAGGGGTGTTCGCGCGAGTCAAGTGAAGTGC 100
DB 2514 CAAGGGGTGCGGGTGGCGGATCACCGTCACTATGCC 2552

```

```

RESULT 14
AAAL4669
ID AAAL4669 standard; DNA; 4818 BP.
XX
XX AAAL4669;
AC
XX
XX 08-AUG-2000 (first entry)
DT
XX
XX Nucleotide sequence of modified FK-520 PKS gene cluster module 8.
DE
XX
XX FK-520; polyketide synthase; PKS; gene cluster; immunosuppressant;
KW immunophilin; FK-506 binding protein; polyketide compound; uveitis;
KW transplant rejection; graft-versus-host disease; alopecia universalis;
KW autoimmune chronic active hepatitis; inflammatory bowel disease;
KW multiple sclerosis; primary biliary cirrhosis; scleroderma;
KW neurite outgrowth; nerve regrowth; Parkinson's disease;
KW Alzheimer's disease; stroke; traumatic spinal cord; brain injury;
KW peripheral neuropathy; ss.
KM
XX
XX Synthetic.
OS Streptomyces hygroscopicus.
OS
XX
XX Key Location/Qualifiers
FH

```

```

FT CDS 3..4817
FT FT /*tag= a
FT FT /note="no termination codon given"
XX
XX WO200020601-A2.
XX
XX 13-APR-2000.
XX
XX
XX 01-OCT-1999; 99WO-US22886.
XX
XX 02-OCT-1998; 98US-0102748.
XX 11-MAR-1999; 99US-0122810.
XX 17-JUN-1999; 99US-0139650.
XX
XX (KOSA-) KOSAN BIOSCIENCES INC.
PA
XX
XX Reeves C, Chu D, Khosla C, Santl D, Wu K;
PI
XX WPI: 2000-317716/27.
XX P-PSDB: AAY84733.
DR
XX
XX New isolated polyketide synthase nucleic acid and polyketide compounds,
PT useful for treating e.g. transplant rejection, uveitis, multiple
PT sclerosis, Alzheimer's disease, Parkinson's disease, stroke, or
PT peripheral neuropathy -
XX
XX Example 2; Page 102-105; 126pp: English.
PS
XX The present sequence represents module 8 of the FK-520 polyketide
CC synthase (PKS) gene cluster, containing the acyltransferase (AT)
CC domain of module 13 of rapamycin. FK-506 is a potent immunosuppressant,
CC and acts through initial formation of an intermediate complex with
CC protein immunophilins known as FK-506 binding proteins. The nucleic
CC acids are used for producing polyketide compounds. The polyketide
CC compounds can be used as immunosuppressants to prevent or treat
CC transplant rejection, graft-versus-host disease or uveitis. They can
CC also be used for treating e.g. alopecia universalis, autoimmune
CC chronic active hepatitis, inflammatory bowel disease, multiple
CC sclerosis, primary biliary cirrhosis, or scleroderma. They
CC also have neurotrophic activity and can be used to promote neurite
CC outgrowth in NGF-treated PC12 cells and in sensory neuronal cultures,
CC and in intact animals, they promote regrowth of damaged facial and
CC sciatic nerves, and repair lesioned serotonin and dopamine neurons in
CC the brain. They can also be used for treating e.g. Parkinson's disease,
CC Alzheimer's disease, stroke, traumatic spinal cord and brain injury, or
CC peripheral neuropathies. They can also be used in agricultural and
CC veterinary applications.
XX
XX Sequence 4818 BP; 742 A; 1982 C; 1476 G; 618 T; 0 other;
SO
Query Match 36.2%; Score 36.6; DB 21; Length 4818;
Best Local Similarity 60.6%; Pred. No. 0.5;
Matches 60; Conservative 0; Mismatches 39; Indels 0; Gaps 0;
OY 2 TCGACGGTCTGCGGGCGGCGAGCGCTGCGAGAGTCTGCGATCTTGCGGCA 61
DB 2505 TCCACCGTATGCGGGCGGCGACCCCGGAAGCGGTGACCATGCTCTACCGCTATGAGCA 2564
OY 62 AAGGGGTGTTCGCGCGAGTCAAGTGAAGTGC 100
DB 2565 CAAGGGGTGCGGGTGGCGGATCACCGTCACTATGCC 2603

```

```

RESULT 15
AAS08693/c
ID AAS08693 standard; DNA; 109519 BP.
XX
XX AAS08693;
AC
XX
XX 26-SEP-2001 (first entry)
DT
XX
XX Micromonospora DNA encoding biosynthetic enzymes for Evernimycin.
DE
XX

```

```
KW Everninomicin; antibiotic; bottle-neck gene; orthomycin;
KM fermentation; ds.
XX
OS Micromonospora carbonacea var. africana.
XX
FH Key Location/Qualifiers
FT CDS complement (132..1382)
FT CDS /*tag= a
FT RBS complement (1389..1394)
FT CDS /*tag= b
FT CDS complement (1490..2611)
FT /*tag= c
FT /*product= "EvdB"
FT RBS complement (2618..2622)
FT /*tag= d
FT CDS complement (2622..3860)
FT /*tag= e
FT /*product= "EvdC"
FT RBS complement (3867..3870)
FT /*tag= f
FT CDS 4143..5312
FT /*tag= g
FT RBS /*product= "EvdD"
FT 4134..4138
FT /*tag= h
FT CDS 5309..6235
FT /*tag= i
FT /*product= "EvdE"
FT CDS 6232..7275
FT /*tag= j
FT /*product= "EvdF"
FT 6226..6229
FT /*tag= k
FT CDS 7272..8327
FT /*tag= l
FT /*product= "EvdG"
FT CDS 8342..9364
FT /*tag= m
FT /*product= "EvdH"
FT RBS 8333..8336
FT /*tag= n
FT CDS complement (9463..10224)
FT /*tag= o
FT /*product= "EvdI"
FT RBS complement (10232..10235)
FT /*tag= p
FT CDS 10424..11176
FT /*tag= q
FT /*product= "EvdJ"
FT CDS 12027..12455
FT /*tag= r
FT /*product= "EvdK"
FT /*partial
FT /*note= "No start codon"
FT CDS complement (12108..13022)
FT /*tag= s
FT /*product= "EvdL"
FT RBS complement (13027..13030)
FT /*tag= t
FT CDS complement (14410..15363)
FT /*tag= u
FT /*product= "EvrA"
FT RBS complement (15369..15373)
FT /*tag= v
FT CDS complement (15380..16414)
FT /*tag= w
FT /*product= "EvrB"
FT CDS complement 16419..17873
FT /*tag= x
FT /*product= "EvrC"
FT CDS complement (17870..18934)
FT /*tag= y
FT /*tag= z
FT CDS complement (19374..20906)
FT /*tag= z
FT /*product= "EvrE"
FT CDS 21064..22542
FT /*tag= aa
FT RBS complement (22542..21056)
FT CDS /*tag= ab
FT 22748..24172
FT /*tag= ac
FT RBS complement (22736..22740)
FT /*tag= ad
FT CDS complement (24177..25223)
FT /*tag= ae
FT /*product= "EvrH"
FT RBS complement (25230..25233)
FT /*tag= af
FT CDS 25550..26626
FT /*tag= ag
FT /*product= "EvrI"
FT CDS 26685..30479
FT /*tag= ah
FT RBS /*product= "EvrJ"
FT 26672..26676
FT /*tag= ai
FT CDS complement (30557..31876)
FT /*tag= aj
FT /*product= "EvrK"
FT RBS complement (31885..31888)
FT /*tag= ak
FT CDS complement (31941..32882)
FT /*tag= al
FT /*product= "EvrL"
FT CDS complement (33167..34405)
FT /*tag= am
FT /*product= "EvrM"
FT RBS complement (34414..34418)
FT /*tag= an
FT CDS complement (34449..35210)
FT /*tag= ao
FT /*product= "EvrN"
FT RBS complement (35219..35221)
FT /*tag= ap
FT CDS complement (35294..36238)
FT /*tag= aq
FT /*product= "EvrO"
FT CDS complement (36235..36963)
FT /*tag= ar
FT /*product= "EvrP"
FT CDS complement (36998..38026)
FT /*tag= as
FT /*product= "EvrQ"
FT CDS complement (38072..38566)
FT /*tag= at
FT /*product= "EvrR"
FT CDS complement (38892..40163)
FT /*tag= au
FT /*product= "EvrS"
FT CDS complement (40216..40890)
FT /*tag= av
FT /*product= "EvrT"
FT RBS complement (40899..40902)
FT /*tag= aw
FT CDS complement (40887..41576)
FT /*tag= ax
FT /*product= "EvrU"
FT CDS complement (41679..42707)
FT /*tag= ay
FT /*product= "EvrV"
FT RBS complement (42714..42717)
FT /*tag= az
```

```
FT CDS complement (42810..43799)
FT /tag- ba
FT /product- "EviM"
FT RBS complement (43807..43811)
FT /tag- bb
FT complement (43799..44866)
FT /tag- bc
FT /product- "EviX"
FT CDS complement (45014..45760)
FT /tag- bd
FT /product- "EviY"
FT RBS complement (45767..45770)
FT /tag- be
FT complement (45962..46714)
FT /tag- bf
FT /product- "EviZ"
FT RBS complement (45952..45956)
FT /tag- bg
FT complement (47156..49234)
FT /tag- bh
FT /product- "EviA"
FT CDS 51627..52715
FT /tag- bi
FT /product- "EviB"
FT RBS 51629..51622
FT /tag- bj
FT 52889..53557
FT /tag- bk
FT /product- "EviC"
FT CDS 53554..54207
FT /tag- bl
FT /product- "EviA"
FT CDS complement (54362..55117)
FT /tag- bm
FT /product- "EviB"
FT RBS complement (55125..55128)
FT /tag- bn
FT complement (55135..56094)
FT /tag- bo
FT /product- "EviC"
FT RBS complement (56100..56103)
FT /tag- bp
FT complement (56184..56813)
FT /tag- bq
FT /product- "EviC2"
FT CDS 56961..58709
```

Query Match 35.2%; Score 35.6; DB 22; Length 109519;

Best Local Similarity 62.2%; Pred.No.1.3; Matches 56; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

```
QY 11 CTGCGCGGAGCCGCGAGGCTGCGAGAGTGCTGCGCATCTTGCGGCAAGGGGgtg 70
DB 25004 CGGCTGGCCCGGCGCGGCGCTGCGGCGGCTGCTGCGCGGCGCGCGCTGCGGCGtg 24945
QY 71 TTCTGCCGTGAGTCAAGGTGACGTCGCC 100
DB 24944 CCCGGCGGCGCGGCGAGCCGTGACGTCGCC 24915
```

Search completed: November 6, 2002, 12:20:05
Job time : 268 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 6, 2002, 11:51:49 ; Search time 443.625 Seconds

(without alignments)
6625.829 Million cell updates/sec

Title: US-09-724-876-2_COPY_33500_33600

Perfect score: 101

Sequence: 1 atcgacggtctgctgcgagcg.....agtcaggtgacgtcgcca 101

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 100%

Listing first 45 summaries

Database : GenEmbl.*

```
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_ov:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_ph:*
24: em_pat:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_in:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rnd:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*
```

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	101	100.0	58733	1	AF217189 Sorangium
2	101	100.0	71989	6	AR172664 Sequence
3	100	99.0	68750	1	AF210843 Sorangium
4	100	99.0	68750	6	AR193029 Sequence
5	100	99.0	68750	6	AR199551 Sequence
6	100	99.0	68750	6	AR199559 Sequence
7	100	99.0	68750	6	AR199567 Sequence
8	100	99.0	68750	6	AR201097 Sequence
9	100	99.0	68750	6	AR208671 Sequence
10	72.8	72.1	10910	1	AX024383 Sequence
11	72.8	72.1	10910	6	AX024276 Sequence
12	58	57.4	66808	1	SAU421825 Stigmatel
13	52.6	52.1	49377	6	188042 Sequence 1
14	52.6	52.1	67523	1	SCU24241 Sorangium c
15	48.4	47.9	49736	1	AF319998 Stigmatel
16	48	47.5	5292	6	AX153793 Sequence
17	48	47.5	5292	6	AX153790 Sequence
18	48	47.5	5292	6	AX153789 Sequence
19	47.8	47.3	77534	1	AF235504 Streptomycin
20	46.2	45.7	42603	1	AF188287 Stigmatel
21	45.2	44.8	33529	6	AR166425 Sequence
22	42.4	42.0	28732	1	AF082100 Streptomycin
23	41.4	41.0	107379	1	SHGCP1R S. hygroscop
24	39.8	39.4	104326	1	AB070940 Streptomycin
25	38.2	37.8	6462	6	AX153794 Sequence
26	38.2	37.8	19791	1	SPK8AD Streptomycin
27	38.2	37.8	77534	1	AF235504 Streptomycin
28	37.6	37.2	10029	1	AE010357 Methanopy
29	37.6	36.6	143786	3	AC024859 Caenorhab
30	37	36.6	298960	2	AC006896 Caenorhab
31	36.8	36.4	104326	1	AB070940 Streptomycin
32	36.6	36.2	27522	1	AB070942 Streptomycin
33	36.6	36.2	107379	1	SHGCP1R S. hygroscop
34	35.6	35.2	1044	6	AX195968 Sequence
35	35.6	35.2	15970	1	AE007061 Mycobacte
36	35.6	35.2	53450	1	MTV018 Mycobacte
37	35.6	35.2	109519	6	AX195929 Sequence
38	35.2	34.9	10670	1	SCARDIGN Saccharothr
39	35	34.7	12381	6	AX006889 Sequence
40	35	34.7	24568	1	AF220951 Streptomycin
41	35	34.7	30690	6	E38020 Avermectin
42	35	34.7	31422	6	E38021 Avermectin
43	35	34.7	47981	1	AF263245 Micromono
44	35	34.7	47981	6	AX112026 Sequence
45	35	34.7	50937	6	AR159871 Sequence

ALIGNMENTS

RESULT 1
AF217189
LOCUS
DEFINITION
AF217189 58733 bp DNA linear BCT 09-JUN-2000
Sorangium cellulosum putative transposase gene, partial cds;
putative transposase gene, complete cds; epothilone biosynthesis
gene cluster, complete sequence; putative membrane protein gene,
complete cds.
ACCESSION
AF217189
VERSION
AF217189.1 GI:7453554
KEYWORDS
POLYANGIUM CELLULOSUM.
ORGANISM
Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
Myxococcales; Sorangineae; Polyangiaceae; Polyangium.
REFERENCE
1 (bases 1 to 58733)

QSTRUMETTLDLKRLPWLGDHRYOGAVVFPAGAYLEMAISSGAEALDGPLOITDVV
LAELAFAGDAVILVOYVTEOPSGRLFOPLASRPGGSHSEFRVHAGALLRVATE
VPAGITLSAVBARLOASMPAAATYALTELMGLOVPAPOAGIATLRRGEGALGRRLP
DAAGSAAYRRLHLPALLDNCFOYVGSIFAGGEGATPWVEVGSLELLOPSELMCHA
RVNHGRQTPROGADFNWVSSGVAVEVSLVAKRLPGGRRREEDWLELEMEP
AAVGTAKNAGRMWLLGGGGLGALRSMLEGGVAHVAAESENISAGVRALAKAF
DGOAPTAVHLGSLDGGELDPGLAGCALAPRASADVSPALDLALVAGCSVMTV
OALAMGFRDAPRLMLTRGAOAVGAGVSTYQAPRLGGRYIAMEHADRCAVYDL
PRDGEGLALLAELADAEVALRGEFRCAIVRROPTREGRREGRREGRREGRRE
IRADSTIYVGTGGLGSLVAGMLERAGRHLYVGRSGASVEDRVAALAEAGAR
VYAKADVADRAQGLRIILREVTTSGLPVGVAHGILLDDGLMOOTPARFEKVAAPK
VOGALHIALHAREAPLEFFVLYASGVGLSGGQGVAAAFELDLAHRRAQOPLPA
LSVDMGLEAEVMAAODRGARLVSRMGSLTPDEGLSALARELSEHRCVQVAVN
PRLWEYLPAAASSRMLSRLYAHASAGSGPAGODDLRLRLAAAPASRSGLEPLR
AOISQVLRLEPGKLEVDALPISIGMNSLMLGELRIRIEMAGITYPATLWYPTVAA
LSGHLARACCAAPESPHTTADSVELEBMSODDLTOLIAKFKALP"

gene
16134..37907
/gene="epod"
/protein_id="AAF62883.1"
/db_xref="GI:7453560"

CDS
16134..37907
/note="type I polyketide synthase; epothilone PKs modules
3, 4, 5, and 6"
/codon_start=1
/transl_table=11
/product="epod"
/protein_id="AAF62883.1"
/db_xref="GI:7453560"

Query Match
Best Local Similarity 100.0%; Score 101; DB 1; Length 58733;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATCGACGGTGTGCGGCGGAGCCGCGCTCGCAGAGTGTGCGCATCTTGCGGC 60
|||||
Db 33500 ATCGACGGTGTGCGGCGGAGCCGCGCTCGCAGAGTGTGCGCATCTTGCGGC 33559
|||||
QY 61 AAAGGGGGTGTCTGCGCGTCAAGGTGAGCGTCGCCA 101
|||||
Db 33560 AAAGGGGGTGTCTGCGCGTCAAGGTGAGCGTCGCCA 33600
|||||

RESULT 2
AR172664 AR172664 71989 bp DNA linear PAT 17-DEC-2001
DEFINITION Sequence 2 from patent US 6303342.
ACCESSION AR172664
VERSION AR172664.1 GI:17912155
KEYWORDS
SOURCE .
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 71989)
AUTHORS Julien,B., Katz,L., Khosla,C. and Tang,L.
TITLE Recombinant methods and materials for producing epothilones C and D
JOURNAL Patent: US 6303342-A 2 16-OCT-2001;
FEATURES
source 1..71989
/organism="unknown"
/protein_id="AAF62883.1"

BASE COUNT 10108 a 23531 c 26617 g 11731 t 2 others

ORIGIN

Query Match
Best Local Similarity 100.0%; Score 101; DB 6; Length 71989;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATCGACGGTGTGCGGCGGAGCCGCGCTCGCAGAGTGTGCGCATCTTGCGGC 60
|||||
Db 33500 ATCGACGGTGTGCGGCGGAGCCGCGCTCGCAGAGTGTGCGCATCTTGCGGC 33559
|||||

QY 61 AAAGGGGGTGTCTGCGCGTCAAGGTGAGCGTCGCCA 101
|||||
Db 33560 AAAGGGGGTGTCTGCGCGTCAAGGTGAGCGTCGCCA 33600
|||||

RESULT 3
AF210843 AF210843 68750 bp DNA linear BCT 21-JAN-2000
LOCUS Sorangium cellulosum strain So ce90 epothilone biosynthesis gene
DEFINITION cluster, complete sequence.
ACCESSION AF210843
VERSION AF210843.1 GI:6724237
KEYWORDS
SOURCE .
ORGANISM
Polyangium cellulosum.
Polyangium cellulosum
Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
Myxococcales; Sorangineae; Polyangiaceae; Polyangium.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Molnar,I., Schnupp,T., Ono,M., Zirkle,R., Milamow,M.,
Nowak-Thompson,B., Engel,N., Toupet,C., Strutzmann,A., Cyr,D.D.,
Goriach,J., Mayo,J.M., Hu,A., Goff,S., Schmidt,J. and Ligon,J.M.
TITLE The biosynthetic gene cluster for the microbule-stabilizing
agents epothilones A and B from Sorangium cellulosum So ce90
JOURNAL Chem. Biol. 7 (2), 97-109 (2000)
MEDLINE 20130945
PUBMED 10662695
REFERENCE 2 (bases 1 to 68750)
AUTHORS Molnar,I.
TITLE Direct Submission
JOURNAL Submitted (03-DEC-1999) Natural Product Genetics, Novartis
Agricultural Research Institute, Inc., 3054 Cornwallis Rd, P.O. Box
12257, Research Triangle Park, NC 27709, USA
FEATURES
source 1..68750
/organism="Polyangium cellulosum"
/strain="So ce90"
/db_xref="taxon:56"
<1..1826
/note="ORF1"
/codon_start=3
/transl_table=11
/product="unknown"
/protein_id="AAF26904.1"
/db_xref="GI:6724238"
/translation="ASLDLAFARATTSARYLDGHRATERHVLAEARGIEDLRALREH
LRIOGSPSHCMGCDLIVLELADHSDPLASISPHARSRLRHPDTSAMLVDPALV
RWLARAGAPRLREYERERARKTOEARRLTLAAAPCFRADLPFRDDANGELPGR
MSPEVAERRLASVATPELACALMLMGAGGAPSGITPAYELPENLDELGELPT
ALAAASAPGTSERLARLGARLFASMEVSSKRSQGNIPDALMERLRTIVRAMGNADN
LSRPERARIAIAEVRRLRROPAPFAAGGLAVAGSSGRSLGVLTDALYSGGND
LYMFOGRASPVLLAGDIDPELAPLSOMFLVHANAGTISKYLTGSPPLIVARN
OARPMSTVHARGFMVNOVADPDERGAPFYVORSTIMEFHPPTPCHEPAGSAFS
LACDEHILWCELSGRGELMHPHRRRCAPRBAVIGENPAAATWYPSITLATHVL
WADPRKAILGVDRKGYEPTVLAETRRPRAVYSEDSDIRALTCQPSRDMVZHHR
SGASTVADYQQLWDRDPMVLRNGLEFFTTNDRLITLARS"

CDS
1..71989
/organism="unknown"
/protein_id="AAF26905.1"

Best Local Similarity 100.0%; Pred. No. 6.5e-14;
Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 TCGACGGTGTCTGCGGGGAGCCGCGACGCTCGACAGAGTGTGGCATCTTCGGGCA 61
|||||
Db 39113 TCGACGGTGTCTGCGGGGAGCCGCGACGCTCGACAGAGTGTGGCATCTTCGGGCA 39172

Qy 62 AAGGGGTGTCTCTGCGGTGAGTCAAGGTGACGTCGCGCA 101
|||||
Db 39173 AAGGGGTGTCTCTGCGGTGAGTCAAGGTGACGTCGCGCA 39212

RESULT 4
ARI93029 68750 bp DNA linear PAT 20-APR-2002
LOCUS Sequence 1 from patent US 6346404.
ACCESSION ARI93029.1 GI:20238994
VERSION ARI93029.1
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and Gorlach,J.
TITLE Genes for the biosynthesis of epoethyllones
JOURNAL Patent: US 6346404-A 1 12-FEB-2002;
FEATURES Location/Qualifiers
source 1..68750
/organism="unknown"

BASE COUNT 9596 a 22456 c 25539 g 11159 t
ORIGIN

Query Match 99.0%; Score 100; DB 6; Length 68750;
Best Local Similarity 100.0%; Pred. No. 6.5e-14;
Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 TCGACGGTGTCTGCGGGGAGCCGCGACGCTCGACAGAGTGTGGCATCTTCGGGCA 61
|||||
Db 39113 TCGACGGTGTCTGCGGGGAGCCGCGACGCTCGACAGAGTGTGGCATCTTCGGGCA 39172

Qy 62 AAGGGGTGTCTCTGCGGTGAGTCAAGGTGACGTCGCGCA 101
|||||
Db 39173 AAGGGGTGTCTCTGCGGTGAGTCAAGGTGACGTCGCGCA 39212

RESULT 5
ARI99551 68750 bp DNA linear PAT 20-APR-2002
LOCUS Sequence 1 from patent US 6355457.
ACCESSION ARI99551
VERSION ARI99551.1 GI:20249625
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and Gorlach,J.
TITLE Genes for the biosynthesis of epoethyllones
JOURNAL Patent: US 6355457-A 1 12-MAR-2002;
FEATURES Location/Qualifiers
source 1..68750
/organism="unknown"

BASE COUNT 9596 a 22456 c 25539 g 11159 t
ORIGIN

Query Match 99.0%; Score 100; DB 6; Length 68750;
Best Local Similarity 100.0%; Pred. No. 6.5e-14;
Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 TCGACGGTGTCTGCGGGGAGCCGCGACGCTCGACAGAGTGTGGCATCTTCGGGCA 61
|||||
Db 39113 TCGACGGTGTCTGCGGGGAGCCGCGACGCTCGACAGAGTGTGGCATCTTCGGGCA 39172

Db 39113 TCGACGGTGTCTGCGGGGAGCCGCGACGCTCGACAGAGTGTGGCATCTTCGGGCA 39172

Qy 62 AAGGGGTGTCTCTGCGGTGAGTCAAGGTGACGTCGCGCA 101
|||||
Db 39173 AAGGGGTGTCTCTGCGGTGAGTCAAGGTGACGTCGCGCA 39212

RESULT 6
ARI99559 68750 bp DNA linear PAT 20-APR-2002
LOCUS Sequence 1 from patent US 6355458.
ACCESSION ARI99559
VERSION ARI99559.1 GI:20249633
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and Gorlach,J.
TITLE Genes for the biosynthesis of epoethyllones
JOURNAL Patent: US 6355458-A 1 12-MAR-2002;
FEATURES Location/Qualifiers
source 1..68750
/organism="unknown"

BASE COUNT 9596 a 22456 c 25539 g 11159 t
ORIGIN

Query Match 99.0%; Score 100; DB 6; Length 68750;
Best Local Similarity 100.0%; Pred. No. 6.5e-14;
Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 TCGACGGTGTCTGCGGGGAGCCGCGACGCTCGACAGAGTGTGGCATCTTCGGGCA 61
|||||
Db 39113 TCGACGGTGTCTGCGGGGAGCCGCGACGCTCGACAGAGTGTGGCATCTTCGGGCA 39172

Qy 62 AAGGGGTGTCTCTGCGGTGAGTCAAGGTGACGTCGCGCA 101
|||||
Db 39173 AAGGGGTGTCTCTGCGGTGAGTCAAGGTGACGTCGCGCA 39212

RESULT 7
ARI99567 68750 bp DNA linear PAT 20-APR-2002
LOCUS Sequence 1 from patent US 6355459.
ACCESSION ARI99567
VERSION ARI99567.1 GI:20249641
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and Gorlach,J.
TITLE Genes for the biosynthesis of epoethyllones
JOURNAL Patent: US 6355459-A 1 12-MAR-2002;
FEATURES Location/Qualifiers
source 1..68750
/organism="unknown"

BASE COUNT 9596 a 22456 c 25539 g 11159 t
ORIGIN

Query Match 99.0%; Score 100; DB 6; Length 68750;
Best Local Similarity 100.0%; Pred. No. 6.5e-14;
Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 TCGACGGTGTCTGCGGGGAGCCGCGACGCTCGACAGAGTGTGGCATCTTCGGGCA 61
|||||
Db 39113 TCGACGGTGTCTGCGGGGAGCCGCGACGCTCGACAGAGTGTGGCATCTTCGGGCA 39172

Qy 62 AAGGGGTGTCTCTGCGGTGAGTCAAGGTGACGTCGCGCA 101
|||||
Db 39173 AAGGGGTGTCTCTGCGGTGAGTCAAGGTGACGTCGCGCA 39212

RESULT	8					
LOCUS	AR201097					
DEFINITION	AR201097	68750 bp	DNA		linear	PAT 20-APR-2002
ACCESSION	Sequence 1	from patent US 6358719.				
VERSION	AR201097					
KEYWORDS	AR201097.1	GI:20251985				
SOURCE	Unknown.					
ORGANISM	Unknown.					
REFERENCE	Unclassified.					
AUTHORS	1 (bases 1 to 68750)					
TITLE	Schnupp,T., Ligon,J.Madison., Molnar,I., Zlickle,R., Cyr,D.Dawn. and Gorlach,J.					
JOURNAL	Genes for the biosynthesis of epoethilones					
FEATURES	Patent: US 6358719-A 1 19-MAR-2002;					
source	Location/Qualifiers					
	1..68750					
	/organism="unknown"					
BASE COUNT	9596 a 22456 c 25539 g 11159 t					
ORIGIN						

[illegible][illegible]

RESULT 10				
AX024383				
LOCUS	AX024383	10910 bp	DNA	linear
				BCT 15-SEP-2000

[illegible]

RESULT	11			
LOCUS	AX024276			
DEFINITION	AX024276	10910 bp	DNA	linear
ACCESSION	Sequence	81 from Patent	DE19846493.	PAT 15-SEP-2000
VERSION	AX024276			
KEYWORDS	AX024276.1	GI:10184550		
SOURCE				
ORGANISM				
	Polyangium cellulorum.			
	Polyangium cellulorum			
	Bacteria; Proteobacteria; delta subdivision; Myxobacteria;			
	Myxococcales; Soranginaeae; Polyangiaceae; Polyangium.			
REFERENCE				
	1 (bases 1 to 10910)			
	Beyer, S. and Mueller, R. J.			
	Patent: DE 19846493-A 81 13-APR-2000.			
JOURNAL				
	BIOTECHNOLOG FORSCHUNG GMBH (DE)			
FEATURES				
source				
	1. 10910			
	/organism="Polyangium cellulorum"			
	/db_xref="taxon:56"			
BASE COUNT	1852 a	3124 c	3613 g	2321 t
ORIGIN				

Query Match	72.1%	Score 72.8	DB 6	Length 10910
Best Local Similarity	83.0%	Pred. No. 1.7e-07		
Matches	83	Conservative	0	Mismatches 17
				Indels 0
				Gaps 0
Qy	2	TCGACGCGTCTGGCGCGGAGCCGGCACGCTCGCAGAGTGTCTGGCCATCTTCGGCGCA	61	
Db	4323	TCGACGCGTCTGGCGCGGAGCCGGCACGCTCTCTGGAGGTGTCTGGCGCGCTGACGGCC	4382	
Qy	62	AAGGGGGTCTTCGCCGTCGAGTCAAGGTGACGTCGCGCA	101	
Db	4383	AAGGGGGTCTTCGCCGCGCAGGTGAAGGTGACGTCGCGCA	4422	

RESULT 12	LOCUS	SAU421825	SAU421825	66608 bp	DNA linear	BC1 16-MAR-2002
DEFINITION	SAU421825	stigmatella aurantiaca OR8, OR7, ORF6, ORF5, OR4, OR3, OR2, OR1, stia gene, stib gene, stic gene, stie gene, stif gene, stid gene, stih gene, stij gene, stik gene, stil gene and				

ACCESSION VERSION KEYWORDS	ORF9
AJ421825	
AJ421825.1	GI:19572309
	acyl-CoA binding protein; cellulase; cytochrome P450 monooxygenase; kinase; methyl transferase; ORF1; ORF2; ORF3; ORF4; ORF5; ORF6; ORF7; ORF9; ORF9: polynucleotide adenyltransferase; potassium channel beta chain; ribosome binding factor; stia gene; Stia gene; stib gene; stib protein; stic gene; stic protein; stid gene; stid protein; stie gene; stie protein; stif gene; stif protein; stig gene; stig protein; stih gene; stih protein; stij gene; stii protein; stik gene; stil gene.
SOURCE	Stigmatella aurantiaca.
ORGANISM	Stigmatella aurantiaca
REFERENCE	Bacteria; Proteobacteria; delta subdivision; Myxobacteria; Myxococcales; Cystobacterineae; Cystobacteraceae; Stigmatella.
AUTHORS	1
TITLE	Galtatzis, N., Silakowski, B., Kunze, B., Nordsiek, G., Blocker, H., Hotle, G. and Muller, R.
JOURNAL	The biosynthesis of the aromatic myxobacterial electron transport inhibitor stigmatellin is directed by a novel type of modular polyketide synthase
REMARK	Online Publication
AUTHORS	J. Biol. Chem., 10.1074/jbc.M11738200
TITLE	2. (bases 1 to 66808)
JOURNAL	Muller, R.
REFERENCE	Direct Submission
KEYWORDS	Submitted (11-DEC-2001) Muller R., MX, GBF, Mascheroderweg 1, 38124 Braunschweig, GERMANY
FEATURES	Location/Qualifiers
source	1..66808
	/organism="Stigmatella aurantiaca"
	/strain="Sg a15"
	/db_xref="taxon:41"
	complement(4978..6231)
	/note="ORF8"
	/codon_start=1
	/transl_table=1
	/product="cytochrome P450 monooxygenase"
	/protein_id="CAD19077.1"
	/db_xref="GI:19572310"
	/translation="MLRSVSYSSRPPSPANPYAANDLARLDDAFADPFPLXRMRKEDPVHMPGSLPTLRWADLDCYTRDTRFFSSDKRAEFGATGDADLEFHHHTSLVFNDPPLHTRVRLIYGVALTPRALSTMEGMLPTLVDRLDGLAVGADLIEDFAAIP EIVYGNLIDPTEERGPRLGWSLA IIGALIEPRLIAEOACGCEAVTEFLDYLRILVQRARPGADATVYLR LIOGESDGERLTETTELHOCVFILNMGHETTTLIGNALETLARFPD EBRLLARAPALIPAVEEVLRYESSNOLGNRRVADPEIDEGVAAPTCTETICGAANNRDPAREDEPEHFDVGRQPNRLARAGAHHCAGNMLARIRALALAFRFDYALTLAPPARRARRRFGFTAMPTRLGSLR"
	6369..7730
	/note="ORF7"
	/codon_start=1
	/transl_table=1
	/product="protein kinase"
	/protein_id="CAD19078.1"
	/db_xref="GI:19572311"
	/translation="MEELISEALPDGSRVGPWRIEGRAGYGTGYAVYRAHRVGRICBR SVALKARIPMDLFEEREAELEFVRANYPRLGCHGTWRGPGGEAHPYVMOWVGGVRLTQWKKRRTARREYLRAVLAOVARALEATHMGGLHDPYCGENVYVSEGHASLMD FCGCTWEGAPLITGCIARPGTKATRSPEALRFHHNHNREDAHIOATIPADVDYALGTTV AYRLCTGIIPPPEATSPGVGDERETQEVKKPPSQVPLVPALFQLLIRMLRDPKPER GRAELAEMKALAEKAGAEADPEPMNHSVPSEITGRPVSYVOREGMPVALVLA VEGLLPGSGSLNTRGRVAPASRMTEGCGTGVADAVALPESASAREPESGLNDMLKVPILPGQGRPPCPRFQISIOGCGWLEVPQSSPCGESYIEMWKACYSVPAAPRRPRS DPO"
	complement(7849..8577)
	/note="ORF6"
	/codon_start=1
	/transl_table=1
	/product="methyl transferase"
	/protein_id="CAD19079.1"
	/db_xref="GI:19572312"
	/translation="MSQNIYADDDPSFEFSEYSQALRSGLAGAPEPALRALLPDVKDL RVYDLAGCGFWCFMWRKNDAAHVLGIDLSEKMLTTRKAKFTSGALITYYOQADLERVL

PCGADPLAFSSAIIHYEELEPLLATYHRCITPGGWVYFSEHIFEMASQPMENIC
 EGHRTWPDVDSYDLGKRTTHMLAEVVAVYHRTIGTTLNALDGGFTLNHIEWPCTDA
 QUSAPHELEAERDRPMLFVSAOR"
 complement (8622. . 8879)
 /note="ORF5"
 /codon_start=1
 /transl_table=11
 /product="acyl-CoA binding protein"
 /protein_id="CAD19080.1"
 /db_xref="GI:19572313"
 /translation="MALEDPPRSQAERYKALKTRPSNDRLLEYLSLFKQATBEDGYGSK
 RRGILIDYKERATYDAMWGKGLGRESAMQYVALYERLRG"
 8961. .10070
 /note="ORF4"
 /codon_start=1
 /transl_table=11
 /product="potassium channel beta chain"
 /protein_id="CAD19081.1"
 /db_xref="GI:19572314"
 /translation="WVYLICRSQSLRWILNGSAHRAVPFANGHARRPRRMHFRHGRS
 GIMAEVSEIGSNMLTHGSOVEEAACVYRALDGIITTFDADVYAGCTRAEEVYGRAL
 GORREDEIFPKYVWPGPGNDKSLRKHILIESIHGSLRLETVEYDLYOAHRYDT
 EPILETWMAFSDVYRQGHKAYHIGSEMTADLRGYSLSKMFQSFOLYSSPOYSMLM
 RYIEAEVYPAQSELGVOGVYSPINQGYLTCTKRCQKPPREGSRATDEKGCASIKRFE
 LNDVLTFRQQLTPVAQERAGLSLAQLAVAWLVQNPVSSALIGASRDEQYKENYKAAQ
 VYLSSEAMHIDTLGPIYVERNPKTKLENSPRKR"
 10428. . 10862
 /note="ORF3"
 /codon_start=1
 /transl_table=11
 /product="ribosome binding factor"
 /protein_id="CAD19082.1"
 /db_xref="GI:19572315"
 /translation="MTSPKORRRAPRGCTESSISSEPAOSLPAHHLRVOSTLFOEVA
 RLFREBLSDPLEGVSVTSFELTPDGRNARIGFTLPREYAGGSAPITAEALVYASGFI
 RQALGLNLRKRVPLHFRFCVGVAPPGPSIPEDGSAEBGES"
 10859. . 12016
 /note="ORF2"
 /codon_start=1
 /transl_table=11
 /protein_id="CAD19083.1"
 /db_xref="GI:19572316"
 /translation="MRPRRIEGAPGAVPILAMARVAPGAVKQOLLAROPYLAEHVA
 AMPDVHAEGVAVGVFAFTERHIVYGAIGDLCGVSAVPDFPAMSLGRVETLRLLA
 AFARAVPGDVAHGRGLSLPGLDIGRLSTHRILOREMERLVPKHGLTGSGNFEL
 DPAADGAMLLIHTSSRGVGAIAHHNRVAYSGESLPRLPDTSEGAACLAPFEL
 ACQFGRANMDTLQRAVEVLGAIGQGVLDSDRVVHNHNAHSHHGRTLVHKKCA
 VGLVAGQGLIPGSMGTASYLVAGGEEQAQSCSHGGRVMTREERERIRPAAHL
 ALRVVYEGRTVALVEEAPAYARDITVELEDEKELVTPLRRLPIAVLNG"
 complement (12082. . 13974)
 /note="ORF1"
 /codon_start=1
 /transl_table=11
 /product="cellulase"
 /protein_id="CAD19084.1"
 /db_xref="GI:19572317"
 /translation="MKNSLILMLIPLVYGTALARPGMANABPHEEGEPENAGLTA
 AOCNRPVLSRGRYIVRCGERFKLSVNMFGASDQLEVYGLDKOKSLDIYAGMKA
 MGFNSVRLPFSNNMLHPDNRKPAQRCELEKISVTCIDSTKPELNTALTEYDAYVA
 ELTRQEVRLIDNNHTTKSMCCGMDNGCSEFMSQALDRMDDWMLVARQYKNGKWA
 ADLRGAEVRDGLDSDPNMGQCNHDMHMAAQMGMLILRTNPDILVEAANVMGMLDGG
 SPQRLKPVQAEVALLRDGLLVYAVHNQYGPNGSGSLGSGKYEADMKPTLYGL
 DQEMFALAAQNAVAPAWMSEFGVGYEQAQNSRAMVSNLADVLEIKDYDMVWAIN
 RAKLONSVOGEDETYGLMYSYPMNMGVRSQDMRLCTGRLITLADGRTGAVDTRRLP
 MVLVLTSTIVDNNSLDSDMHSGRAKISCPGYRMGVSTANFTTLCTNAGAVPAQ
 OGTGVYQITTEVSPURYPGDMANOSIKFECPGYFVAGVSTANPFLLEAGLICEQN
 TGGILNLSAANFWGDDORASSSGGWSVQYKGCACADNOYLLIGLAHRRSGIADYPS
 VYLCSN"
 14345. . 21466
 /gene="stia"
 14345. . 21466
 /gene="stia"
 /function="PKS"

Query Match 52.1%; Score 52.6; DB 1; Length 67523;
Best local Similarity 70.7%; Pred. No. 0.0072;
Matches 70; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

Qy 2 TCACAGCTGCTGGCGGCGGACGCCGCTGCAGAGCTGCTGCAGATCTTGCGGCA 61
Db 16698 TCACAGCTGCTGGCGGCGGACGCCGCTGCAGAGCTGCTGCAGATCTTGCGGCA 16757

Qy 62 AAGGGGCTGTTGCGGTGAGTCAAGGCTGAGAGTGGCC 100
Db 16758 AAGAGCGCTTTGTCGTCGAGTGGCAGTGCAGCTGCC 16796

RESULT 15
AF319998/c
LOCUS 49736 bp DNA linear BCT 28-MAY-2001
DEFINITION Stigmatella aurantiaca myxalamid biosynthetic gene cluster,
complete sequence.
ACCESSION AF319998
VERSION AF319998.1 GI:14210834
KEYWORDS
SOURCE Stigmatella aurantiaca.
ORGANISM Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
Stigmatella aurantiaca.
REFERENCE 1 (bases 1 to 49736)
AUTHORS Silakowski, B., Nordsiek, G., Kunze, B., Bloeker, H. and Mueller, R.
TITLE Novel features in a combined polypeptide synthase/non-ribosomal
peptide synthetase: the myxalamid biosynthetic gene cluster of the
myxobacterium Stigmatella aurantiaca Sgal5
Chem. Biol. 8 (1), 59-69 (2001)

JOURNAL 2 (bases 1 to 49736)
MEDLINE Silakowski, B., Nordsiek, G., Bloeker, H. and Mueller, R.
PUBMED 11182319
REFERENCE Direct Submission
AUTHORS Submitted (07-NOV-2000) MX, GBF, Mascheroderweg 1, Braunschweig
JOURNAL 38124, Germany

FEATURES
source Location/Qualifiers
1..49736
/organism="Stigmatella aurantiaca"
/db_xref="taxon:41"
complement(2..1306)
/note="ORF1"
/codon_start=1
/transl_table=11
/product="putative aspartokinase"
/protein_id="AAK57182.1"
/db_xref="GI:14210834"
/translation="MRPISAROGAGTSLALIVOKYGTGVSQDPERIKYNAKRCIA
AOKAGHVVVVVAMSGETNRLLKLYSOTTERPNFEDOVVAVTGEQVSGIALAIO
TQKRAVSFLGQVRIITDSTSKARIKSIDERIYDALKKNIYVAVGFCQDDEHN
VTTLRGSDTTAVALAALAKADACEIYTDVGVTYTDNNVPAARKERIAIEEELK
LASVGAQVLOIRSEVFAMKYKVLWKSFTDDPGLVCEEDAAEDVYVSGIAVDN
EAKIARIGVPIQIRSEVFAMKYKVLWKSFTDDPGLVCEEDAAEDVYVSGIAVDN
DVKKIASKVKAEGVETDQVYKIVGVMRNSGVAKMFTVLAGAVNNQIMSTS
ETKVCVSHSNTELAIVNSLHAFGLDRPEAGAVSEVYSLK"
1363..2895
/note="ORF2"
/codon_start=1
/transl_table=11
/product="putative histidine ammonium lyase"
/protein_id="AAK57183.1"
/db_xref="GI:14210836"
/translation="MSRPLNIDGTLKLEETILQVARIHTVVELAPAAARVKAARDL

gene complement(3275..7822)
/gene="mxaa"
CDS /gene="mxaa"
complement(3275..7822)
/note="non-ribosomal peptide synthetase; contains a
condensing, adenylation, and PCP reductase domains"
/codon_start=1
/transl_table=11
/product="Mxas"
/protein_id="AAK57184.1"
/db_xref="GI:14210837"
/translation="MSVRELLFTTLAKGVRLRLEGGKLVQVEASKAVLTPELRAELSAN
KADILAFLRKSSREFTSELGOPLARPEVIRPLSSGOERLWFLDRLSGTSGVANHG
LBARGLDPRKARRSIDELVRRHEVLRTHFEPEVSSPROVITYSPPTGVALTVELPGL
TESQREELTRSDLAQDFDLAKGLRLRVYLAGPDQFLATYTKHITTDGWSLG
VFVRELSTMLYASFVRGQAALPPVSMQFDSALRRERAMLAGESGARERAYWMLKGLG
PELOLPVDHASTSTSHRGATVPALASPLASDRLRELASREGCSFTVYFAFASALH
RSGQVDFGVGVIANRGSVPADLIGFINTALRLCDLSGEPTFSOMLARAIVLEA
LDYQALPFESEVQVGTGASRDGDLNPLVRACPTLESTIPATLDPGTSNPLGADGDS
VEGVAKFEELSLTAASEKLAGLELREYRPPASVVERVWGHQVYLLSIVNHPVL
SKPLTLAEERGLNDNRGPLYDPAOLMELVQAOKERTQAAVAVSGQKTLTYAE
LRRNQALNHLRLRGLIOKEERVNGLCEVETEDIVGLALILAKGAVVPLDPAVERA
LALLIEDAQVAVLLTQOQLVPELPATQAVVLDMDPITGESPSTNEBTRTAPARA
VLYTSGSTGKGVNIEHRNVAELILMAVSFSEKLAGTASTSICFDSLVSFIET
PLCCGKAVYVAKNALELPELPARETILNTVYVSMGALIRSGVSSAIVLEA
LAGALVDIOYLDHYRDVNVNLGYSSETTYSPTFVNNGOFTIRGPNQVAYLDS
NREPPETVPGSVYIGMGVANGYIGRPELTARERVRSPFGGSGEARLRYTDLRLM
PQGLEIYRMDHGVKLRGEFLEIGELVLESHGIRDAVNVGVRGADQOLAVY
VGRGEKAPPAELRDYLRKLEPEYVPLFGLDALPLPNKVRBALPAERVHSG
PAKEHVAPPTPEESLAALIMRWOLVPEVQIGANDFEELGSHLLRYLVILARSAGA
DILRLALDAPTLLEMARAVEAKGTSLPAHVYVMEADVLDAEILGKRLSPYNG
ALRTILTGATGFLGATFLLELCRRTPDAIYCYLVSKRTQOEGMNRIRKNSIAME
ALAPRIVPVKDIGPLALCSKEFORLSEEDALYHNGALVNLTPIPESRRAAVLG
TREILRLATRTIKPLHYVTVSVLPLGKRAIRDEPLEGSSLVGGVGAOSKVAEK
LVREASRGLPTIILRGVTHGSTRGTGAMNDLVCRTLKCGRGVRLVPDALLDT
PDVYSSAIVDSMRPESIGQTYHLVNPQVADMMNVMRAFGRGLRVPDQWLS
LGSAASSDSELDGLIMFLQOVPPEDRSVGPRMNVCDSDTLKALGIGTSCPSVDS
LISTYLSLVHNGFLKAPBVR"
complement(7819..11520)
/gene="mxabi"
complement(7819..11520)
/note="PKS: ketosynthase; acyltransferase"
/codon_start=1
/transl_table=11
/product="Mxabi"
/protein_id="AAK57185.1"
/db_xref="GI:14210838"
/translation="MRAGAGHPILGVSFTTSVOKENRPNWSALSAKPSYLADHRVG
GAVVYGAAYELMALSAARELSGEKAYELVDSREGLLPEGCGRVOMALSEGGD
ALSPQSSOKASTERNAPRGMAVYGGRIPEGSPSTPESPAITRESEVLAQA
AHYEELASGVAVYPAFGQVODIMGRRAELRHVLPBLRAASAVYIHPALDAC
QVAAVAPDEDTPAGPAVPAVALAHLRHEDPGTEVCHVRLHRSBEARAYVEFDL
QVLRASGRILVALRLRQOOLDAPAABHAKGDLFPALSMRHSAAVOVAPARGVSGR
WLLINDGSELADTVESOLBERIEYIRVLDIAGGGRVAVDTSPQALVTRVYHSGS
GAPRGVYHITGLPEPEYDPIWGCGGVLIHQDALSRVLSAPSPRLVTRVYHSGS
KSASDVVQASLWGLRTTLAHEHSELKTRVDYSAMGTGDGALMLRELVADDEEEL
SLRADGRHVRIVRGSPGRAPSEAVVAPGGGRFHLERADQGLERFAAGARRPPGEL
ELEVEVALGGPGEGTGERMEAGEGCGSCGVRALIEGVKNAVVAAGEERIALVAGL
GSHVSPAGCTAIRPSSSSHGCAALATVYPAWVLLHGLERGRVWVYGAAGSL
GRAATIKLRAGAEEFAVATTEORASALREIGADVOMPSDPSRVLDVATGGAGVE
LATVAPGAVELEPRLPTLSEGRVYDLDAASGASATGGAANYAVYVODODARRRPER
FASLMOEVKAAREAGCLPSAADAKALMLKDKARVYVPAADARLRADGYTLVYTGGL
GGLGLAVAKMWEAGHLLVYGRDTVLSPEQEVVYALAEAGARVAVRADVSDRAQ

```

gene
CDS
IARLSEIAGGAPLRGVIIHAGVLDGVLLOQTVERFRRYMAKVLGVMNLHVLTRE
APLDFEVLYSASAMFGAPGOGNYVAGNAFDLALHRRRALGGLGLSINNGPEFEVL
AAAGANGERBIAAGSDSLTPAEGNTILGRLLDGDVTOMAVMPLRLKRWYFPRAS
SPMLSELVPADAGCEVEARDVALLETLRTRTPDEAKAALAEQVREOLARVLRDSAR
IDEPALQSGFLDSLMGLELRNLRLVSGVSLPSLIWKHPTLAALCAHLGELVMDRT
IAETLAQAAGAPADVASADKEVFL"
complement (11487..14303)
/gene="maxb2"
complement (11487..14303)
/note="PKS; dehydratase; enoylreductase; ketoreductase;
ACP"
/codon_start=1
/transl_table=11
/product="Mxab2"
/protein_id="AAK57186.1"
/db_xref="GI:14210839"
/translation="MSKOEKPAEQPSALQRAALIVERMQARLDABERATEPIAVIGL
GCRFPGGAVDSSSYWKILRDGVDALEVPESRMDVPGHFSTRGVPGKMTGRGFLD
DVEHDEFEFGISPREAASLDPOORVLEVAEMALEMAGLADPOLAGSKTGVGMWS
SDYMARLKENADATPDGMATNGCYSFVGRSVYVGLGQPCPVPTACSSSLVAIH
LASESLRAGESNLAAGVNLISPEPTICLSMOALASDRCKTDAADGVIRGEG
CGVLVLRISDAQRDGDSILALIRGSAVNHGASGGLTPNGPQQAQVQORALDNARI
APALVGYIEAHGTGTPLDPIELRALGAVLGKGRPADRPFISGVTNIGHLEPAAGI
AGVYKTLISQHKIEIPHLHERPNNPMDRIIPARVVERVPMWAEGRRIAGVSSEF
GLSGINAHVYLEEAPASTAPVAEEDKAOLVYLSAKTPRALSLEAMSSFLKNEEG
GSLADISYTSALSTPREHRLALAASKOELIROLEPAAKSHAGLSGSRGTGLCO
KITVFPGGSSOMLGWGORLTAENVTFREAIOWCHEAISRITDMSLRDILTPNGQPR
MNDIDIVLOPTLPAQVLAAVMWLGLPEHAAVHGSHGEVAAAHVAGALSLDEARII
CERSKILRRVSGOGAMAAYVDLSREOAEEIRGAEDRIAIVSNGRSTVLSGDPAAL
EVLRLERREVFCRWKVDVAHSPOKPLRODLDRMAAIRPTGSAVPLYSTVGAP
IDGRLESAFVWVRNLRQVLEADAVORLIDOGVNLFTLSPHPTLVPFDAMLREGDT
SGRLVYPSLIRDOEPPSLSSVGAFTVYIANTGSGYIPTVAGLRPLTRGSASG
TGMSCDOLRRRAGSCALARKEVTS"
complement (14300..29878)
/gene="maxc"
complement (14300..29878)
/note="PKS; modules 1,2 and 3 contain KS, AT, DH, KR, and
ACP regions"
/codon_start=1
/transl_table=11
/product="Mxac"
/protein_id="AAK57187.1"
/db_xref="GI:14210840"
/translation="MANRNDSENELSPAKRMVALERKQTRLNABEGANEPIAYGM
ACRPQADNVEAFWRLLQGVDAVREIPLERWPKDEYADIGDAMRWGFLDSYSDFD
PGFEGISPREAVRMDPOORLLEVAWEALBDAGLDVXLGSRSGVFIGACNDYHGM
OVERPETGDAFSATGVAASVLSGRSLYLFNLOGPSLVVDTRAGSSIVSLHACOSLRN
RECMALAGVNLILSPSVLVSQIALSPDGRSAFASANGFTRGEGCIYVLR
LSDALADGDIILATIRGSAINODCKSTGLTTPNYLSOQALIRQALLESSGLKAEVSTY
EAHGTGTPLDPILEAELRETYGVPRPDGSCVCGISVNTVGHLSAAGIAGIMKVL
AMRHQSLPPLHLKQVNPRIQLBGSALTPTRLTPMKASGQPRRAGVSFSGISCTNAH
VILEEAPAPAKDALPQREVLPLSARSRLVALAERYAERLREGGSLSDCYTA
SAHRTHTHRLAVLADREHTAERLRAFARGEOPSGAMSGKARGRKTIYFVYPOG
AORAGMRALMESEPVRDALVYDKALOPHIGMSVLDETAADORSLEAIDVNPQV
LRAVEVALPLAHSWGIEDPDAVHGSHGEVYAAHVAACALSLEDAVLTICHSKLMRTL
GGKATAMVVELSVEEAEAIRPNNGKIMLGNGLPSSOVLSGEPEAVEALGKR
GVFTRWKMDVPSHTPLVEVISRLEGRRIARTPRADIPICSTVGEFELDGALDAA
HMGNNLNNPVRFAAAQHLASDHGTVEVSPHVLPAVEOTMSAGNTEGVAVLASLR
RNEDERTHTLSIGALYSGRVAMNVRVYEGKGLVPLPVTYPMQREHFWMEPVLASGR
ASRTSSGHPLIGERIDLSAGEVSRIMVOLGPKHPLIAEHRYQVAPLPAAMADM
ALSAKOTLGEFVSLEDFVSALISPEDEERYQLMLSTERPGASATPLFSAPVRE
TPSDELTWALHABGSLRHDASVVERKAPALDELRLARCEBAEVAEHYOTLKEIGLE
YGSFVRVRELIRAKGALGRVQLEGLTAEERYSLHAPLDDCFQVLAALAEGLDS
EGGYVVALSRLLVLRPEAALSHVIRIVSAGASNTIECDVTLILDGRRVAAEVR
RLVCKRTAAASAPADEVDNWIYSAOFEPOROAARETSSRTAGRMVLPADARIGOS
LASGLARAGGCYVVEAGPYERLEADRYRINPSKLEDPQALILKEAFSDSPCKGMVH
LYGIDVANSNTEPESFRALSLSTSVYVILQAMAGRMRTPLMLVYTTGVQVYIAGS
EQPESPAAPLVGLAAVITHEHPELRCAHYDLAQSPSEETASLTTELIDGGAEDRIA
LRGAVRVVARLGRIGISGRKAAEPREFADGTYLITGGIGGLEAVAKWFIORGARNLA
LTGCAPESEAKQVLELRANGAEVRYRADVTSAADVARLIGSIDSELPRLRGVHA
AGVADGVLTQIDPEPLRAVLAIPVQGMNMLHVQTOGRPLELTFVMSAASMSGPGO

```

```

Query Match          47.9%; Score 48.4; DB 1; Length 49736;
Best Local Similarity 68.4%; Pred. No. 0.071;
Matches 67; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

OY 2 TCGACGGTGTCTGCGCGGAGCGGACGCGTCCGACAGAGTGTGCGATCCTTGGCGCA 61
    |||||
Db 45824 TCGACGGTGTCTGCGCGGAGCGGACGCGTCCGACAGATTCGATCTTGCACCGC 45765
    |||||

OY 62 AAGGGGTGTCTGCGCGTCCGACGTCAGGTGAGCGTCCG 99
    |||||
Db 45764 CCGGAGTGTCTGCGCGGCGGTGATATCAACGTCC 45727
    |||||

```

Search completed: November 6, 2002, 13:22:46
 Job time : 592.625 secs

```

GNYAANAEMDVLAHHRRAAGLPGLSYNMGMFADVGLAAVATNRGRLARGVGNMP
AQALEAIGRLIVQAOAVVAPLDLRQWREPLYSAAOSPLSYMOKKYSTPARKGS
ARELLAAABPQRKGIIEGYLRDELGILIRNEPSRIEDPOFGDGLGDSLTJGLRBR
IESGLKTLSTLTIYATPTVOALTLYLKGLEPYLEABARTVAPTOTEPREVSQVAP

```


RESULT 4
LOCUS BG368755
DEFINITION HVSME10020107f Hordeum vulgare 20 DAP spike EST library HVC-DNA0010 (20 DAP) Hordeum vulgare cDNA clone HVSME10020107f, mRNA sequence.
ACCESSION BG368755
VERSION BG368755.1 GI:13257856
KEYWORDS EST.
SOURCE Hordeum vulgare.
ORGANISM Hordeum vulgare.
REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidaeae; Triticeae; Hordeum.
AUTHORS Wing, R., Close, T.J., Kleinbols, A., Wise, R., Begum, D., Frisch, D., Yu, X., Henry, D., Palmer, M., Rambo, T., Simmons, J., Choi, D.W., Fenton, R.D., Close, S.J., Oates, R. and Main, D.
TITLE Development of a genetically and physically anchored EST resource for barley genomics: Morex 20 DAP spike cDNA library
JOURNAL Unpublished (2001)
COMMENT Contact: Wing RA
 Clemson University Genomics Institute
 100 Jordan Hall, Clemson, SC 29634, USA
 Tel: 864 656 7288
 Fax: 864 656 4293
 Email: rwing@clemson.edu
 Total hg bases = 546
 Seq primer: AATTACCTTCCTCAAGG
 High quality sequence stop: 717.
FEATURES
 source
 1..819
 /organism="Hordeum vulgare"
 /cultivar="Morex"
 /db_xref="taxon:4513"
 /clone="HVSME10020107f"
 /clone_1fb="Hordeum vulgare 20 DAP spike EST library
 HVC-DNA0010 (20 DAP)"
 /tissue_type="20 DAP spike"
 /lab_host="SOLR"
 /note="Vector: lambdaZAP; Site_1: EcoRI; Site_2: XhoI; Plants were grown in the greenhouse at the University of California, Riverside (Fenton, SJ Close, TJ Close). Whole spikes with awns trimmed were collected at 20 DAP (Fenton). Total RNA was prepared, poly(A) RNA was purified, one primary unamplified cDNA library was made, and 1 million pfu were in vivo excised to give pBluescript SK(-) cDNA phagemids in the TJ Close lab at the University of California, Riverside (Choi). Phagemids were plated and picked at the Clemson University Genomics Institute (CUGI) (Begum, Palmer, Frisch, Atkins and Wing). Plasmid DNA preparations, DNA sequencing and sequence analysis were performed at CUGI (Wing, Yu, Frisch, Henry, Simmons, Oates, Rambo, Main). The sequence has been trimmed to remove vector sequence and contains a minimum of 100 bases of phred value 20 or above. For more details on library preparation and sequence analysis see <http://www.genome.clemson.edu/projects/barley>. To order this clone see <http://www.genome.clemson.edu/orders> Also see Close TJ, Wing R, Kleinbols A, Wise R (2001) Genetically and physically anchored EST resources for barley genomics. Barley Genetics Newsletter 31:25-30. (<http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html>)."
BASE COUNT 128 a 255 c 308 g 128 t
ORIGIN
 Query Match 32.7%; Score 33; DB 12; Length 819;
 Best Local Similarity 65.8%; Pred. No. 60;
 Matches 48; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

RESULT 5
LOCUS BF482518/c
DEFINITION WHE1795_E07_113S wheat pre-anthesis spike cDNA library Triticum aestivum cDNA clone WHE1795_E07_113, mRNA sequence.
ACCESSION BF482518
VERSION BF482518.1 GI:11565819
KEYWORDS EST.
SOURCE bread wheat.
ORGANISM Triticum aestivum
REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidaeae; Triticeae; Triticum.
AUTHORS Anderson, O.D., Chao, S., Choi, D.W., Close, T.J., Fenton, R.D., Han, P.S., Hsia, C.C., Kang, Y., Lazo, G.R., Miller, R., Rausch, C.J., Seaton, C.L. and Tong, J.C.
TITLE The structure and function of the expressed portion of the wheat genomes - Pre-anthesis spike cDNA library
JOURNAL Unpublished (2000)
COMMENT Contact: Olin Anderson
 US Department of Agriculture, Agriculture Research Service, Pacific West Area, Western Regional Research Center
 800 Buchanan Street, Albany, CA 94710, USA
 Tel: 5105595773
 Fax: 5105595818
 Email: oanderson@pw.usda.gov
 Sequence have been trimmed to remove vector sequence and low quality sequence with phred score less than 20
 Seq primer: Stratagene SK primer.
FEATURES
 source
 1..607
 /organism="Triticum aestivum"
 /cultivar="Chinese Spring"
 /db_xref="taxon:4565"
 /clone="WHE1795_E07_113"
 /clone_1fb="Wheat pre-anthesis spike cDNA library"
 /tissue_type="Spike before anthesis"
 /dev_stage="Adult plant"
 /lab_host="E. coli SOLR"
 /note="Vector: Lambda Uni-ZAP XR, excised phagemid; Site_1: EcoRI; Site_2: XhoI; Plants were grown in the greenhouse. Whole spike with awns trimmed, white, green and yellow anther were collected and total RNA, and poly(A) RNA were prepared, a cDNA library was made, and the cDNA clones were in vivo excised to give pBluescript phagemids in the TJ Close lab (Choi, Close, Fenton) at the University of California, Riverside. Plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (all other authors)."
BASE COUNT 108 a 196 c 128 g 175 t
ORIGIN
 Query Match 32.5%; Score 32.8; DB 12; Length 607;
 Best Local Similarity 64.5%; Pred. No. 65;
 Matches 49; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

RESULT 6
LOCUS 2
DEFINITION GCGGCTGCGAGCGCTGCGAGCGAGCGCTTCGCGCCGCGAGCGAGCG 61
ACCESSION 414 GCGGCGCGCGAGCGCTGCGAGCGAGCGATGCTGATGCGCGAGCGAGCGAGCG 473
VERSION 1
KEYWORDS EST.
SOURCE Hordeum vulgare.
ORGANISM Hordeum vulgare.
REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidaeae; Triticeae; Hordeum.
AUTHORS Wing, R., Close, T.J., Kleinbols, A., Wise, R., Begum, D., Frisch, D., Yu, X., Henry, D., Palmer, M., Rambo, T., Simmons, J., Choi, D.W., Fenton, R.D., Close, S.J., Oates, R. and Main, D.
TITLE Development of a genetically and physically anchored EST resource for barley genomics: Morex 20 DAP spike cDNA library
JOURNAL Unpublished (2001)
COMMENT Contact: Wing RA
 Clemson University Genomics Institute
 100 Jordan Hall, Clemson, SC 29634, USA
 Tel: 864 656 7288
 Fax: 864 656 4293
 Email: rwing@clemson.edu
 Total hg bases = 546
 Seq primer: AATTACCTTCCTCAAGG
 High quality sequence stop: 717.
FEATURES
 source
 1..819
 /organism="Hordeum vulgare"
 /cultivar="Morex"
 /db_xref="taxon:4513"
 /clone="HVSME10020107f"
 /clone_1fb="Hordeum vulgare 20 DAP spike EST library
 HVC-DNA0010 (20 DAP)"
 /tissue_type="20 DAP spike"
 /lab_host="SOLR"
 /note="Vector: lambdaZAP; Site_1: EcoRI; Site_2: XhoI; Plants were grown in the greenhouse at the University of California, Riverside (Fenton, SJ Close, TJ Close). Whole spikes with awns trimmed were collected at 20 DAP (Fenton). Total RNA was prepared, poly(A) RNA was purified, one primary unamplified cDNA library was made, and 1 million pfu were in vivo excised to give pBluescript SK(-) cDNA phagemids in the TJ Close lab at the University of California, Riverside (Choi). Phagemids were plated and picked at the Clemson University Genomics Institute (CUGI) (Begum, Palmer, Frisch, Atkins and Wing). Plasmid DNA preparations, DNA sequencing and sequence analysis were performed at CUGI (Wing, Yu, Frisch, Henry, Simmons, Oates, Rambo, Main). The sequence has been trimmed to remove vector sequence and contains a minimum of 100 bases of phred value 20 or above. For more details on library preparation and sequence analysis see <http://www.genome.clemson.edu/projects/barley>. To order this clone see <http://www.genome.clemson.edu/orders> Also see Close TJ, Wing R, Kleinbols A, Wise R (2001) Genetically and physically anchored EST resources for barley genomics. Barley Genetics Newsletter 31:25-30. (<http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html>)."
BASE COUNT 128 a 255 c 308 g 128 t
ORIGIN
 Query Match 32.7%; Score 33; DB 12; Length 819;
 Best Local Similarity 65.8%; Pred. No. 60;
 Matches 48; Conservative 0; Mismatches 25; Indels 0; Gaps 0;


```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="6p21.3: HLA Class I region"
/clone="25R probe"
/clone_lib="Human"
/Note="Vector: pMOS Blue"

BASE COUNT      269 a 450 c 406 g 375 t
ORIGIN

Query Match      31.7%; Score 32; DB 17; Length 1500;
Best Local Similarity 58.3%; Pred. No. 1.2e+02;
Matches 56; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

OY      2  GCCGCGTCGAGCGCTGCCGCGAGCGAGCGTGCAGAGCTTCTCGGCCGACGCGAGC 61
         |||||  |||  |||||  |||  |||||  |||  |||||  |||  |||||
Db      630 GCCGCGTCGACTGCTCCCGGGCCCTCGGAGCGTGCAGAGCTGCTGGAGAGGTGAGCA 571
OY      62  GCGTGGGTGAGCGAGCGAGCGCTGCCCATGCTCTGC 97
         |||||  |||  |||  |||  |||  |||  |||  |||  |||
Db      570 GGCATGGGAGAGTGAAGTGGGGGCTGACCCAGC 535

RESULT 9
BQ241000/c      650 bp  mRNA  linear  EST 03-MAY-2002
LOCUS           TAE05010C09R TAE05 Triticum aestivum cDNA TAE05010C09R, mRNA
DEFINITION      sequence.
ACCESSION       BQ241000
VERSION         BQ241000.1 GI:20436876
KEYWORDS        EST.
SOURCE          bread wheat.
ORGANISM        Triticum aestivum
                Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
                Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae
                ; Triticeae; Triticum.
                1 (bases 1 to 650)
REFERENCE
  AUTHORS       Cloutier,S.
  TITLE         Wheat functional genomics - Glenlea developing seeds cDNA libraries
  JOURNAL       Unpublished (2002)
  COMMENT       Contact: Dr. Sylvie Cloutier
                Cereal Research Centre, Agriculture and Agri-Food Canada
                195 Dufour Rd, Winnipeg, MB, Canada R3T 2M9
                Tel: (204) 983-2340
                Fax: (204) 983-4604
                Email: scloutier@em.agr.ca
                was cloned directionally, not all sequences generated with reverse
                primer were from the 5' end (same with forward primer and 3' end).
                Average insert size is >2.0 kb
                Plate: 010 row: C column: 09
                Seq primer: M13 Reverse.

FEATURES
  source
    1..650
    /organism="Triticum aestivum"
    /cultivar="Glenlea"
    /db_xref="taxon:4565"
    /clone="TAE05010C09R"
    /clone_lib="rap05"
    /tissue_type="developing seeds"
    /dev_stage="5 days after anthesis"
    /lab_host="E. coli DH10B"
    /note="Vector: pSPORT-P (Invitrogen Technologies); Site_1:
    NotI; Site_2: MluI; mRNA obtained from wheat seeds of
    cultivar Glenlea 5 days post-anthesis"

BASE COUNT      111 a 235 c 162 g 141 t 1 others
ORIGIN

Query Match      31.5%; Score 31.8; DB 14; Length 650;
Best Local Similarity 64.0%; Pred. No. 1.2e+02;
Matches 48; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

OY      16  CTGGCTCCCGAGCGAGCGTCTTCGCGCCGACGCGCGTGGGTGAGC 75
         |||||  |||  |||||  |||  |||||  |||  |||||  |||  |||||
Db      277 CTGGCGCCGACCGCGGAGCGGAGCGAGCGCGCGGCGAAGCGCGCGCGAGC 218

```

```

OY      76  GAAGCTGCGCGCANG 90
         |  |||||  |||
Db      217 GCGCGCTGCGCGANG 203

RESULT 10
A1861213/c      369 bp  mRNA  linear  EST 19-JUL-1999
LOCUS           603011D02.x1 603 - stressed root cDNA library from Wang/Bonnett lab
DEFINITION      Zea mays cDNA, mRNA sequence.
ACCESSION       A1861213
VERSION         A1861213.1 GI:5525215
KEYWORDS        EST.
SOURCE          Zea mays.
ORGANISM        Zea mays.
                Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
                Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
                clade; Panicoidae; Andropogoneae; Zea.
                1 (bases 1 to 369)
REFERENCE
  AUTHORS       Walbot,V
  TITLE         Maize ESTs from various cDNA libraries sequenced at Stanford
                University
  JOURNAL       Unpublished (1999)
  COMMENT       Contact: Walbot V
                Department of Biological Sciences
                Stanford University
                855 California Ave, Palo Alto, CA 94304, USA
                Tel: 650 723 2227
                Fax: 650 725 8221
                Email: walbot@stanford.edu
                Plate: 603011 row: D column: 02.

FEATURES
  source
    1..369
    /organism="Zea mays"
    /cultivar="B73"
    /db_xref="taxon:4577"
    /clone_lib="603 - stressed root cDNA library from
    Wang/Bonnett lab"
    /tissue_type="seedling"
    /dev_stage="salt stress"
    /lab_host="E. coli XL Gold"
    /note="Organ: root; Vector: pBluescriptII SK(+); XR:
    Seedling stressed root cDNA library from Wang/Bonnett lab"

BASE COUNT      74 a 127 c 99 g 69 t
ORIGIN

Query Match      31.3%; Score 31.6; DB 9; Length 369;
Best Local Similarity 58.5%; Pred. No. 1.2e+02;
Matches 55; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

OY      8  TGCAGAGCCTGCTCCGAGCGAGCGTGCAGAGCTTCTCGCCGACGCGCGCTGC 67
         |||  |||||  |  |||  |||||  |||  |||||  |||  |||||
Db      295 TGGCTGCTCTCCAGAGCGAGCGAGCGATCTCGAGCGCGGACATCTTCGTGG 236
OY      68  GGTGAGCGAGCGCGCGCATGCTGCTGCA 101
         |  |||  |||||  |||  |||||  |||  |||||  |||
Db      235 GCGGCGCGTGGCAGCGACTGCACTGGCCGA 202

RESULT 11
A1947509/c      495 bp  mRNA  linear  EST 19-AUG-1999
LOCUS           603022E10.x1 603 - stressed root cDNA library from Wang/Bonnett lab
DEFINITION      Zea mays cDNA, mRNA sequence.
ACCESSION       A1947509
VERSION         A1947509.1 GI:5739819
KEYWORDS        EST.
SOURCE          Zea mays.
ORGANISM        Zea mays.
                Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
                Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
                clade; Panicoidae; Andropogoneae; Zea.

```



```

RESULT 14
BI004847      223 bp  mRNA      linear  EST 13-JUN-2001
DEFINITION   M4-HN0056-150301-001-g10 HN0056 Homo sapiens cDNA, mRNA sequence.
ACCESSION    BI004847
VERSION      BI004847.1  GI:14408921
KEYWORDS     EST.
SOURCE       human.
ORGANISM     Homo sapiens
              Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE    1 (bases 1 to 223)
              Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
              Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
              Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
              Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
              ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
              Simpson,A.J.
              Shotgun sequencing of the human transcriptome with ORF expressed
              sequence tags
              Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
JOURNAL      20202663
MEDLINE
COMMENT      Contact: Simpson A.J.G.
              Laboratory of Cancer Genetics
              Ludwig Institute for Cancer Research
              Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
              Brazil
              Tel: +55-11-2704922
              Fax: +55-11-2707001
              Email: asimpson@ludwig.org.br
              This sequence was derived from the FAPESP/LICR Human Cancer Genome
              Project. This entry can be seen in the following URL
              (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=MR4&ct=MR4-HN0056-
              150301-001-g10&ts=2001-03-15&tl=1)
              Seq primer: puc 18 forward
              High quality sequence start: 38
              High quality sequence stop: 223.
FEATURES
  source
    1..223
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone_id="HN0056"
    /dev_stage="Adult"
    /note="Organ: head,normal; Vector: puc18; Site_1: SmaI;
    Site_2: SmaI; A mini-library was made by cloning products
    derived from ORBESTES PCR (U.S. Letters Patent application
    No. 196,716 - Ludwig Institute for Cancer Research)
    profiles into the puc 18 vector. Reverse transcription of
    tissue mRNA and cDNA amplification were performed under
    low stringency conditions."
BASE COUNT   41 a      72 c      79 g      30 t      1 others
ORIGIN
Query Match      31.1%; Score 31.4; DB 13; Length 223;
Best Local Similarity 59.6%; Pred. No. 1.3e+02;
Matches 53; Conservative 0; Mismatches 36; Indels 0; Gaps 0;
QY      2  GCCGGCTGCAGAGCCTGCTCCGACGAGCGAGCTTCGCGCCGACGCCGACG 61
      11 111111111111111111111111111111111111111111111111111
Db      85  GCACGGTCAGGAGGTGGCGCGCGCTGCCATTCGACGACCTTCGGGCGCGCTCTGCG 144
      11 111111111111111111111111111111111111111111111111111
QY      62  GCGTGGGTGGAGCGAAGCGCTGCGGCATG 90
      11 111111111111111111111111111111111111111111111111111
Db      145  CCGCCAGGCTGACCGAATGCTGCGCGGTG 173
      11 111111111111111111111111111111111111111111111111111
RESULT 15
BG556725      410 bp  mRNA      linear  EST 10-APR-2001
LOCUS        BG556725
DEFINITION   EML_40_G08_g1_A002 Embryo 1 (EML) sorghum bicolor cDNA, mRNA
              sequence.

```

```

ACCESSION    BG556725
VERSION      BG556725.1  GI:13585723
KEYWORDS     EST.
SOURCE       Sorghum.
ORGANISM     Sorghum bicolor
              Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
              Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
              clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE    1 (bases 1 to 410)
              Reid,S.P., Cordomier-Pratt,M.-M., Gingle,A. and Pratt,L.H.
              An EST database from Sorghum: developing embryos
              Unpublished (2000)
              Contact: Cordomier-Pratt MM
              Laboratory for Genomics and Bioinformatics
              The University of Georgia, Department of Plant Biology
              Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
              Tel: 706 542 1860
              Fax: 706 583 0210
              Email: mpratt@uga.edu
              Sequences have been trimmed to exclude PolyA, vector and regions
              below phred quality 16. The threshold for highest quality sequence
              is 20.
              Seq primer: polyTmix
              High quality sequence start: 46
              High quality sequence stop: 405
              POLYA=yes.
FEATURES
  source
    1..410
    /organism="Sorghum bicolor"
    /db_xref="taxon:4558"
    /clone_id="Embryo 1 (EML)"
    /note="Organ: Embryos germinated for 24 hr; Vector:
    pBluescript II from lambda zap II; Site_1: XhoI; Site_2:
    EcoRI; The library was made from poly-A RNA in the cloning
    vector lambda ZAP II. Clones to be sequenced were
    prepared by mass excision."
BASE COUNT   69 a      120 c      164 g      57 t
ORIGIN
Query Match      31.1%; Score 31.4; DB 12; Length 410;
Best Local Similarity 59.6%; Pred. No. 1.4e+02;
Matches 53; Conservative 0; Mismatches 36; Indels 0; Gaps 0;
QY      8  TTGGAGGCGCTGCGCCGACGAGCGTGCAGAGCTTCGCGCGACGCGCGCGTGG 67
      111111111111111111111111111111111111111111111111111
Db      197  TTGGCTGCTTACCAAGACAGCGACGCGCATGCTGAGGCGCGGACATCTTCTGTG 256
      111111111111111111111111111111111111111111111111111
QY      68  GGTGAGCGAAGGCTGCGCCATGCTCTCG 96
      111111111111111111111111111111111111111111111111111
Db      257  GCGGCGGCTGCGCAGCGATCGCACCTG 285
      111111111111111111111111111111111111111111111111111
Search completed: November 6, 2002, 15:52:31
Job time : 1201.62 secs

```



```
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600,438US1
; CURRENT APPLICATION NUMBER: US/09/861,289
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 36
; LENGTH: 4041
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-861-289-36
```

```
Query Match          50.9%; Score 51.4; DB 10; Length 4041;
Best Local Similarity 69.3%; Pred. No. 1.9e-06;
Matches 70; Conservative 0; Mismatches 31; Indels 0; Gaps 0;
```

```
QY 1 AGCGGCTGAGAGGCTGCTCCCGACGAGCGGTGCAAGAGCTTCTCGGCCGAGCGGAC 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 739 AGCAGCCAGCAGGCGCATGGCCCGACGCGCGGACCAAGGGCTTCCGCTCGGCCGCGAC 798

QY 61 GCGGTGGGTGAGCGAAGGCTGCGCCATGCTCTCTCTCAA 101
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 799 GGCCTCGCTGGGCGAGGGCGCTGCGCTACTCTCTCTCGA 839
```

```
RESULT 3
US-09-861-289-32
; Sequence 32, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600,438US1
; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 32
; LENGTH: 11220
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-861-289-32
```

```
Query Match          50.9%; Score 51.4; DB 10; Length 11220;
Best Local Similarity 69.3%; Pred. No. 2e-06;
Matches 70; Conservative 0; Mismatches 31; Indels 0; Gaps 0;
```

```
QY 1 AGCGGCTGAGAGGCTGCTCCCGACGAGCGGTGCAAGAGCTTCTCGGCCGAGCGGAC 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 742 AGCCGTCAGCGGCGGCTGGCGCGGAGCGCGGCTCGAATCTCTCTCGAGCGCGGAC 801

QY 61 GCGGTGGGTGAGCGAAGGCTGCGCCATGCTCTCTCTCAA 101
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 802 GGCACCACTGTGTCGAGGCGCTGCGCTCTCTCTCTCGA 842
```

```
RESULT 4
US-09-861-289-5
; Sequence 5, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
```

```
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600,438US1
; CURRENT APPLICATION NUMBER: US/09/861,289
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 36778
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-861-289-5
```

```
Query Match          50.9%; Score 51.4; DB 10; Length 36778;
Best Local Similarity 69.3%; Pred. No. 2.1e-06;
Matches 70; Conservative 0; Mismatches 31; Indels 0; Gaps 0;
```

```
QY 1 AGCGGCTGAGAGGCTGCTCCCGACGAGCGGTGCAAGAGCTTCTCGGCCGAGCGGAC 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 16429 AGCCGTCAGCGGCGGCTGGCGCGGAGCGCGGCTCGAATCTCTCTCTCGAGCGCGAC 16488

QY 61 GCGGTGGGTGAGCGAAGGCTGCGCCATGCTCTCTCTCAA 101
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 16489 GGCACCACTGTGTCGAGGCGCTGCGCTCTCTCTCTCGA 16529
```

```
RESULT 5
US-09-861-289-30
; Sequence 30, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600,438US1
; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 30
; LENGTH: 13842
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-861-289-30
```

```
Query Match          44.6%; Score 45; DB 10; Length 13842;
Best Local Similarity 65.3%; Pred. No. 0.00013;
Matches 66; Conservative 0; Mismatches 35; Indels 0; Gaps 0;
```

```
QY 1 AGCGGCTGAGAGGCTGCTCCCGACGAGCGGTGCAAGAGCTTCTCGGCCGAGCGGAC 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3961 AGCCGTCAGCGGCGGCTGGCGCGGAGCGCGGCTCGAATCTCTCTCGAGCGCGGAC 4020

QY 61 GCGGTGGGTGAGCGAAGGCTGCGCCATGCTCTCTCTCAA 101
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 4021 GGCACCACTGTGTCGAGGCGCTGCGCTCTCTCTCTCGA 4061
```

```
RESULT 6
US-09-861-289-34
; Sequence 34, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
```


US-09-985-675-9

Query Match 28.1%; Score 28.6; DB 10; Length 4364;
Best Local Similarity 59.0%; Pred. No. 5.2;
Matches 49; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 18 GGTCCCGACGAGAGCTTCTCGCCGACGCCGAGCGCGTGGGTGAGCGA 77
DB 744 GGACTCCGCGACGAGCGCGGTGGCGCGGACGCGGAGCGCGGAGCGCC 685
QY 78 AGGCTGCCCATGCTCTGCTCA 100
DB 684 CGGCGCCCATGCTCGCCTCA 662

RESULT 10

US-09-924-859A-2/c
Sequence 2, Application US/09924859A
Patent No. US20020137113A1
GENERAL INFORMATION:
APPLICANT: Godowski, Paul J.
APPLICANT: Mark, Melanie R.
APPLICANT: Sadick, Michael D.
APPLICANT: Shelton, David L.
APPLICANT: Wong, Mai Lee Tan
TITLE OF INVENTION: KINASE RECEPTOR ACTIVATION ASSAY
FILE REFERENCE: P0854C1P2C1
CURRENT APPLICATION NUMBER: US/09/924,859A
PRIOR FILING DATE: 1999-10-13
PRIOR APPLICATION NUMBER: US/09/417,381
NUMBER OF SEQ ID NOS: 11
SEQ ID NO 2
LENGTH: 2742
TYPE: DNA
ORGANISM: Homo Sapien
US-09-924-859A-2

Query Match 28.1%; Score 28.4; DB 10; Length 2742;
Best Local Similarity 60.3%; Pred. No. 5.8;
Matches 47; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 23 CCGACGAGCGGTGCAAGCTTCTCGCCGACGCCGAGCGCGTGGAGCGAGGCT 82
DB 85 CCGCCGACGACGACCGCGCGGTGGCGGCGGACGCGGAGCGCGCGCC 26
QY 83 GCGCATGCTCTGCTCA 100
DB 25 GCCCATGCTCGCCTCA 8

RESULT 11

US-09-223-490-1/c
Sequence 1, Application US/09223490
Patent No. US20020147325A1
GENERAL INFORMATION:
APPLICANT: Godowski, Paul J.
APPLICANT: Mark, Melanie R.
APPLICANT: Scadden, David T.
APPLICANT: Baker, Kevin P.
APPLICANT: Baron, Will F.
TITLE OF INVENTION: Protein Tyrosine Kinases
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/223,490
FILING DATE:

CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/170,558
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Hasak, Janet E.
REGISTRATION NUMBER: 28,616
REFERENCE/DOCKET NUMBER: 854C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-1896
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3611 bases
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-223-490-1

Query Match 28.1%; Score 28.4; DB 10; Length 3611;
Best Local Similarity 60.3%; Pred. No. 5.8;
Matches 47; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 23 CCGACGAGCGGTGCAAGCTTCTCGCCGACGCCGAGCGCGTGGGTGAGCGAGGCT 82
DB 91 CCGCCGACGACGACCGCGCGGTGGCGGCGGACGCGGAGCGCGCGCC 32
QY 83 GCGCATGCTCTGCTCA 100
DB 31 GCCCATGCTCGCCTCA 14

RESULT 12

US-09-815-242-4093
Sequence 4093, Application US/09815242
Patent No. US2002061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
PRIOR FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 4093

LENGTH: 1395
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-815-242-4093

Query Match 27.7%; Score 28; DB 10; Length 1395;
Best Local Similarity 55.0%; Pred. No. 7.3;
Matches 55; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 2 GCCGGCTGCGAGCCCTGCTCCGACGAGCGTGCAGAGCTTCTCGCCGACGCG 61
DB 482 GCCGCAACCCGACCCGACGCGCGCGGCGGATCGCGCGAGCATCTCCGACG 541
QY 62 GCGTGGGTGGAGCGAGCGTGGCGCATGCTCTGCTCAA 101
DB 542 GCGTGTCTTGGCGAATGCGACCGCATGCTCGGATCAA 581

RESULT 13
US-09-006-298-26
Sequence 26, Application US/09006298
Patent No. US2002008224A1
GENERAL INFORMATION:
APPLICANT: Jolly, Douglas J.
APPLICANT: Moore, Margaret D.
APPLICANT: Chada, Sunil
TITLE OF INVENTION: NON-IMMUNOGENIC PRODRUGS AND SELECTABLE
MARKERS FOR USE IN GENE THERAPY
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED AND BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/006,298
FILING DATE: 13-JAN-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Mcmasters, David D.
REGISTRATION NUMBER: 33,963
REFERENCE/DOCKET NUMBER: 930049, 459
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 2410 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 17..2398
US-09-006-298-26

Query Match 27.7%; Score 28; DB 10; Length 2410;
Best Local Similarity 56.5%; Pred. No. 7.4;
Matches 52; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 5 GCGTCGAGAGCGCTGCGCCGACGAGCGTTCGCGCGCGACCGCGAGCGCG 64
DB 1746 GCGTGGCCGTCCTCCAGAAAGAGTGGCTCAAGACCCCTCAGTCAGTCCGCTGTG 1805
QY 65 TGGGTGGAGCGAGGCTGCGCCCATGCTCTG 96

DB 1806 TGTGTGCGAGGAAGCTTCTCCTGCACAG 1837

RESULT 14
US-09-880-107-3714
Sequence 3714, Application US/09880107
Patent No. US20020142981A1
GENERAL INFORMATION:
APPLICANT: Horne, Darci T.
APPLICANT: Vockley, Joseph G.
APPLICANT: Scherf, Uwe
APPLICANT: Gene Logic, Inc.
TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
FILE REFERENCE: 44921-5028-WO
CURRENT APPLICATION NUMBER: US/09/880,107
CURRENT FILING DATE: 2001-06-14
PRIOR APPLICATION NUMBER: US 60/211,379
PRIOR FILING DATE: 2000-06-14
PRIOR APPLICATION NUMBER: US 60/237,054
PRIOR FILING DATE: 2000-10-02
NUMBER OF SEQ ID NOS: 3950
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 3714
LENGTH: 4180
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: Genbank Accession No. US20020142981A1 X17094
US-09-880-107-3714

Query Match 27.7%; Score 28; DB 10; Length 4180;
Best Local Similarity 56.5%; Pred. No. 7.6;
Matches 52; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 5 GCGTCGAGAGCGCTGCGCCGACGAGCGTTCGCGCGCGACCGCGAGCGCG 64
DB 1946 GCGTGGCCGTCCTCCAGAAAGAGTGGCTCAAGACCCCTCAGTCAGTCCGCTGTG 2005
QY 65 TGGGTGGAGCGAGGCTGCGCCCATGCTCTG 96
DB 2006 TGTGTGCGAGGAAGCTTCTCCTGCACAG 2037

RESULT 15
US-09-799-875-10
Sequence 10, Application US/09799875
Patent No. US20020034780A1
GENERAL INFORMATION:
APPLICANT: Meyers, Rachel
APPLICANT: Kapeller-libermann, Rosana
TITLE OF INVENTION: No. US20020034780A1 Human Protein Kinases and Uses
FILE REFERENCE: 35800/209996
CURRENT APPLICATION NUMBER: US/09/799,875
CURRENT FILING DATE: 2001-03-06
PRIOR APPLICATION NUMBER: 60/182,059
PRIOR FILING DATE: 2000-02-11
PRIOR APPLICATION NUMBER: 09/659,287
PRIOR FILING DATE: 2000-09-12
NUMBER OF SEQ ID NOS: 32
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 10
LENGTH: 2162
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (76)...(1818)
US-09-799-875-10

Query Match 27.3%; Score 27.6; DB 10; Length 2162;
Best Local Similarity 58.5%; Pred. No. 9.6;

	Matches	48;	Conservative	0;	Mismatches	34;	Indels	0;	Gaps	0;
Qy	5	GGCTGGAGGCTTGGCTCCCGAGGAGGTGCAAGACTTCTGGCCGACCGAGCGGCG	64							
Db	54	GGCAGCGGGGGCGGGCCACCATGGGGGTGGACAGCGCGCTGGCCCGGCGCTGCAGCT	113							
Qy	65	TGGGGTGGAGCGAAGGCTGCGC	86							
Db	114	GGGTCGAGCGCTGCTGCGC	135							

Search completed: November 6, 2002, 20:31:24
 Job time : 42.125 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 6, 2002, 11:51:54 : Search time 33.5 Seconds
(without alignments)
924.608 Million cell updates/sec

Title: US-09-724-876-2_COPY_32070_32170

Perfect score: 101
Sequence: 1 agccgctgcgagcctgcgc.....tgcgcctctctctgtctcaaa 101

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapept 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 682724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA: *
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq: *
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq: *
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq: *
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq: *
5: /cgn2_6/ptodata/2/ina/PCITUS_COMB.seq: *
6: /cgn2_6/ptodata/2/ina/backfile1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	101	100.0	68750	3	US-09-335-409-1
2	101	100.0	68750	4	US-09-568-102-1
3	101	100.0	68750	4	US-09-567-969-1
4	101	100.0	68750	4	US-09-568-480-1
5	101	100.0	68750	4	US-09-568-486-1
6	101	100.0	68750	4	US-09-568-472-1
7	101	100.0	68750	4	US-09-567-899-1
8	101	100.0	71989	4	US-09-443-501A-2
9	56.2	55.6	33529	4	US-09-144-085-3
10	54.6	54.1	751	3	US-09-010-809-4
11	54.4	53.9	15872	4	US-09-105-537-1
12	54.4	53.9	43280	2	US-08-804-227C-1
13	53.2	52.7	80161	3	US-09-036-987A-1
14	53.2	52.7	80161	4	US-09-370-700-1
15	52.6	52.1	28958	1	US-08-258-261B-6
16	52.6	52.1	28958	1	US-08-456-837-6
17	52.6	52.1	28958	1	US-08-457-342-6
18	52.6	52.1	28958	1	US-08-457-646A-6
19	52.6	52.1	28958	1	US-08-458-076A-6
20	52.6	52.1	28958	1	US-08-764-233A-4
21	52.6	52.1	28958	1	US-08-457-335A-6
22	52.6	52.1	28958	1	US-08-729-214-6
23	52.6	52.1	28958	3	US-09-028-934-6
24	52.6	52.1	49377	1	US-08-764-233A-1
25	51.4	50.9	1462	4	US-09-434-288-4
26	51.4	50.9	4041	4	US-09-105-537-36
27	51.4	50.9	11230	4	US-09-105-537-32

28	51.4	50.9	36778	4	US-09-105-537-5	Sequence 5, Appl1
29	51.4	50.9	38506	3	US-09-320-878-19	Sequence 19, Appl1
30	50.2	49.7	44377	2	US-08-804-227C-7	Sequence 7, Appl1
31	50.2	49.7	44377	2	US-08-804-198-1	Sequence 1, Appl1
32	50	49.5	11219	1	US-07-642-734C-1	Sequence 1, Appl1
33	50	49.5	11219	3	US-08-439-009A-1	Sequence 1, Appl1
34	49.8	49.3	20235	1	US-07-642-734C-3	Sequence 3, Appl1
35	49.8	49.3	20235	3	US-08-439-009A-3	Sequence 3, Appl1
36	47.6	47.1	440365	4	US-09-103-840A-2	Sequence 2, Appl1
37	47.6	47.1	4411529	4	US-09-103-840A-1	Sequence 1, Appl1
38	45.8	45.3	13987	2	US-08-804-227C-13	Sequence 13, Appl1
39	45.2	44.8	50937	4	US-09-428-517-1	Sequence 1, Appl1
40	45	44.6	13842	4	US-09-105-537-30	Sequence 30, Appl1
41	43.8	43.4	423	4	US-09-144-085-5	Sequence 5, Appl1
42	41.8	41.4	1419	4	US-09-434-288-2	Sequence 2, Appl1
43	40.4	40.0	4689	4	US-09-105-537-34	Sequence 34, Appl1
44	39.6	39.2	333	4	US-09-060-756-318	Sequence 318, App
45	39.6	39.2	1434	4	US-09-434-288-3	Sequence 3, Appl1

ALIGNMENTS

```
RESULT 1
US-09-335-409-1
: Sequence 1, Application US/09335409
: Patent No. 6121029
: GENERAL INFORMATION:
: APPLICANT: Schupp, Thomas
: APPLICANT: Ligon, James
: APPLICANT: Molnar, Istvan
: APPLICANT: Zirkle, Ross
: APPLICANT: Cyr, Devon
: APPLICANT: Goerlach, Joern
: TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
: FILE REFERENCE: 4-30582A
: CURRENT APPLICATION NUMBER: US/09/335,409
: CURRENT FILING DATE: 1999-06-17
: NUMBER OF SEQ ID NOS: 30
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 1
: LENGTH: 68750
: TYPE: DNA
: ORGANISM: Sorangium cellulosum
US-09-335-409-1

Query Match      100.0%; Score 101; DB 3; Length 68750;
Best Local Similarity 100.0%; Pred. No. 2, 6e-19;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGCCGCTGCGAGCCTGCTCCGACGAGCGTGCAGAGACCTTCGGCCGACCGCAG
Db 37682 AGCCGCTGCGAGCCTGCTCCGACGAGCGTGCAGAGACCTTCGGCCGACCGCAG 37741
QY 61 GCGCTGGGCTGAGCGAAGGCTGGCCATGCTCTGCTCAAA 101
Db 37742 GCGCTGGGCTGAGCGAAGGCTGGCCATGCTCTGCTCAAA 37782

RESULT 2
US-09-568-102-1
: Sequence 1, Application US/09568102
: Patent No. 6346404
: GENERAL INFORMATION:
: APPLICANT: Schupp, Thomas
: APPLICANT: Ligon, James
: APPLICANT: Molnar, Istvan
: APPLICANT: Zirkle, Ross
: APPLICANT: Cyr, Devon
: APPLICANT: Goerlach, Joern
: TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
: FILE REFERENCE: 4-30582A
: CURRENT APPLICATION NUMBER: US/09/568,102
```

```
; CURRENT FILING DATE: 2000-05-10
; FILE REFERENCE: 4-30582A
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-102-1
```

```
Query Match          100.0%; Score 101; DB 4; Length 68750;
Best Local Similarity 100.0%; Pred. No. 2.6e-19;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 AGCCGGCTGCGAGGCTGCTGCCGACGAGCGGTGCAAGAGCTTCTCGCCGCGCGGAC 60
Db 37682 AGCCGGCTGCGAGGCTGCTGCCGACGAGCGGTGCAAGAGCTTCTCGCCGCGCGGAC 37741
```

```
QY 61 GCGGTGGGTGAGCGAGGCTGCGCCATGCTCTCTCTCA 101
Db 37742 GCGGTGGGTGAGCGAGGCTGCGCCATGCTCTCTCTCA 37782
```

```
RESULT 3
US-09-567-969-1
; Sequence 1, Application US/09567969
; Patent No. 6355457
; GENERAL INFORMATION:
```

```
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/567,969
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-567-969-1
```

```
Query Match          100.0%; Score 101; DB 4; Length 68750;
Best Local Similarity 100.0%; Pred. No. 2.6e-19;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 AGCCGGCTGCGAGGCTGCTGCCGACGAGCGGTGCAAGAGCTTCTCGCCGCGCGGAC 60
Db 37682 AGCCGGCTGCGAGGCTGCTGCCGACGAGCGGTGCAAGAGCTTCTCGCCGCGCGGAC 37741
```

```
QY 61 GCGGTGGGTGAGCGAGGCTGCGCCATGCTCTCTCTCA 101
Db 37742 GCGGTGGGTGAGCGAGGCTGCGCCATGCTCTCTCTCA 37782
```

```
RESULT 4
US-09-568-480-1
; Sequence 1, Application US/09568480
; Patent No. 6355458
; GENERAL INFORMATION:
```

```
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
```

```
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,480
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-480-1
```

```
Query Match          100.0%; Score 101; DB 4; Length 68750;
Best Local Similarity 100.0%; Pred. No. 2.6e-19;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 AGCCGGCTGCGAGGCTGCTGCCGACGAGCGGTGCAAGAGCTTCTCGCCGCGCGGAC 60
Db 37682 AGCCGGCTGCGAGGCTGCTGCCGACGAGCGGTGCAAGAGCTTCTCGCCGCGCGGAC 37741
```

```
QY 61 GCGGTGGGTGAGCGAGGCTGCGCCATGCTCTCTCTCA 101
Db 37742 GCGGTGGGTGAGCGAGGCTGCGCCATGCTCTCTCTCA 37782
```

```
RESULT 5
US-09-568-486-1
; Sequence 1, Application US/09568486
; Patent No. 6355459
; GENERAL INFORMATION:
```

```
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,486
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-486-1
```

```
Query Match          100.0%; Score 101; DB 4; Length 68750;
Best Local Similarity 100.0%; Pred. No. 2.6e-19;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 AGCCGGCTGCGAGGCTGCTGCCGACGAGCGGTGCAAGAGCTTCTCGCCGCGCGGAC 60
Db 37682 AGCCGGCTGCGAGGCTGCTGCCGACGAGCGGTGCAAGAGCTTCTCGCCGCGCGGAC 37741
```

```
QY 61 GCGGTGGGTGAGCGAGGCTGCGCCATGCTCTCTCTCA 101
Db 37742 GCGGTGGGTGAGCGAGGCTGCGCCATGCTCTCTCTCA 37782
```

```
RESULT 6
US-09-568-472-1
; Sequence 1, Application US/09568472
; Patent No. 6358719
; GENERAL INFORMATION:
```

```
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
```

```

: APPLICANT: Zirkle, Ross
: APPLICANT: Cyr, Devon
: APPLICANT: Goerlach, Joern
: TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
: FILE REFERENCE: 4-30582A
: CURRENT APPLICATION NUMBER: US/09/568,472
: CURRENT FILING DATE: 2000-05-10
: PRIOR APPLICATION NUMBER: 09/335,409
: PRIOR FILING DATE: 1999-06-17
: NUMBER OF SEQ ID NOS: 30
: SOFTWARE: Patentln Ver. 2.0
: SEQ ID NO 1
: LENGTH: 68750
: TYPE: DNA
: ORGANISM: Sorangium cellulosum
US-09-568-472-1

```

```

Query Match 100.0%; Score 101; DB 4; Length 68750;
Best Local Similarity 100.0%; Pred. No. 2.6e-19;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 AGCCGGCTGCGAGGCTGCTCCGACGAGGTCAGAGCTTCTCGCCGACCGAC 60
Db 37682 AGCCGGCTGCGAGGCTGCTCCGACGAGGTCAGAGCTTCTCGCCGACCGAC 37741
Qy 61 GCGGTGGGTGAGGAGGCTGCGCCATGCTCTCTCTCAA 101
Db 37742 GCGGTGGGTGAGGAGGCTGCGCCATGCTCTCTCTCAA 37782

```

```

RESULT 7
US-09-567-899-1
: Sequence 1, Application US/09567899
: Patent No. 6383787
: GENERAL INFORMATION:
: APPLICANT: Schupp, Thomas
: APPLICANT: Ligou, James
: APPLICANT: Molnar, Istvan
: APPLICANT: Zirkle, Ross
: APPLICANT: Cyr, Devon
: APPLICANT: Goerlach, Joern
: TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
: FILE REFERENCE: 4-30582A
: CURRENT APPLICATION NUMBER: US/09/567,899
: CURRENT FILING DATE: 2000-05-10
: PRIOR APPLICATION NUMBER: 09/335,409
: PRIOR FILING DATE: 1999-06-17
: NUMBER OF SEQ ID NOS: 30
: SOFTWARE: Patentln Ver. 2.0
: SEQ ID NO 1
: LENGTH: 68750
: TYPE: DNA
: ORGANISM: Sorangium cellulosum
US-09-567-899-1

```

```

Query Match 100.0%; Score 101; DB 4; Length 68750;
Best Local Similarity 100.0%; Pred. No. 2.6e-19;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 AGCCGGCTGCGAGGCTGCTCCGACGAGGTCAGAGCTTCTCGCCGACCGAC 60
Db 37682 AGCCGGCTGCGAGGCTGCTCCGACGAGGTCAGAGCTTCTCGCCGACCGAC 37741
Qy 61 GCGGTGGGTGAGGAGGCTGCGCCATGCTCTCTCTCAA 101
Db 37742 GCGGTGGGTGAGGAGGCTGCGCCATGCTCTCTCTCAA 37782

```

```

RESULT 8
US-09-443-501A-2
: Sequence 2, Application US/09443501A
: Patent No. 6303342
: GENERAL INFORMATION:

```

```

: APPLICANT: Kosan Biosciences, Inc.
: APPLICANT: Julien, Bryan
: APPLICANT: Katz, Leonard
: APPLICANT: Khosla, Chaitan
: APPLICANT: Tang, Li
: APPLICANT: Ziermann, Rainer
: TITLE OF INVENTION: Recombinant Methods and Materials for Producing
: FILE REFERENCE: 30062-20031.00
: CURRENT APPLICATION NUMBER: US/09/443,501A
: CURRENT FILING DATE: 1999-11-19
: PRIOR APPLICATION NUMBER: US 60/130,560
: PRIOR FILING DATE: 1999-04-22
: PRIOR APPLICATION NUMBER: US 60/122,620
: PRIOR FILING DATE: 1999-03-03
: PRIOR APPLICATION NUMBER: US 60/119,386
: PRIOR FILING DATE: 1999-02-10
: PRIOR APPLICATION NUMBER: US 60/109,401
: PRIOR FILING DATE: 1998-11-20
: NUMBER OF SEQ ID NOS: 22
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 2
: LENGTH: 71989
: TYPE: DNA
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Synthetic construct
US-09-443-501A-2

```

```

Query Match 100.0%; Score 101; DB 4; Length 71989;
Best Local Similarity 100.0%; Pred. No. 2.6e-19;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 AGCCGGCTGCGAGGCTGCTCCGACGAGGTCAGAGCTTCTCGCCGACCGAC 60
Db 32070 AGCCGGCTGCGAGGCTGCTCCGACGAGGTCAGAGCTTCTCGCCGACCGAC 32129
Qy 61 GCGGTGGGTGAGGAGGCTGCGCCATGCTCTCTCTCAA 101
Db 32130 GCGGTGGGTGAGGAGGCTGCGCCATGCTCTCTCTCAA 32170

```

```

RESULT 9
US-09-144-085-3
: Sequence 3, Application US/09144085
: Patent No. 6280999
: GENERAL INFORMATION:
: APPLICANT: Gustafsson, Claes
: APPLICANT: Betlach, Mary C.
: APPLICANT: Ashley, Gary
: APPLICANT: Julien, Bryan
: APPLICANT: Ziermann, Rainer
: TITLE OF INVENTION: SORANGIUM POLYKETIDE SYNTHASES AND ENCODING DNA
: FILE REFERENCE: 30062-20020.20
: CURRENT APPLICATION NUMBER: US/09/144,085
: CURRENT FILING DATE: 1998-08-31
: EARLIER APPLICATION NUMBER: 09/010,809
: PRIOR FILING DATE: 1998-01-22
: NUMBER OF SEQ ID NOS: 8
: SOFTWARE: Patentln Ver. 2.0
: SEQ ID NO 3
: LENGTH: 33529
: TYPE: DNA
: ORGANISM: Sorangium cellulosum
US-09-144-085-3

```

```

Query Match 55.6%; Score 56.2; DB 4; Length 33529;
Best Local Similarity 72.3%; Pred. No. 3.7e-07;
Matches 73; Conservative 0; Mismatches 28; Indels 0; Gaps 0;
Qy 1 AGCCGGCTGCGAGGCTGCTCCGACGAGGTCAGAGCTTCTCGCCGACCGAC 60
Db 32130 GCGGTGGGTGAGGAGGCTGCGCCATGCTCTCTCTCAA 32170

```

Db 376 AGCCGGCTCAAGGGGATGGCCCGGACGGCCGCTGCAGAGACTTCTCTGGCGAGCTGAC 435
QY 61 GGCGTGGGTGAGAGAGCTGCGCATCTCTCTGCTCA 101
Db 436 GGCGTCACTGTGTCGAGAGGGGTGCGGATCTGTCTGTA 476

RESULT 10

US-09-010-809-4
; Sequence 4, Application US/09010809B
; Patent No. 6080601
; GENERAL INFORMATION:
; APPLICANT: Gustafsson, Claes
; APPLICANT: Bellach, Mary C.
; TITLE OF INVENTION: Epithelione Polyketide Synthases and Encoding DNA
; FILE REFERENCE: 30062-20020.00
; CURRENT APPLICATION NUMBER: US/09/010,809B
; CURRENT FILING DATE: 1998-01-22
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 751
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-010-809-4

Query Match Best Local Similarity 71.3%; Score 54.6; DB 3; Length 751;
Matches 72; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 1 AGCCGGCTGAGAGCTGCTCCGACGAGCTGCAAGAGCTTCGCGCCGACGCGAC 60
Db 130 AGTGTGCGCGGGGCTTGGCGAGAGCGTGGTGCAGAGAGCTTCTGCTGCGAGGCTGAC 189
QY 61 GGCGTGGGTGAGAGAGCTGCGCATCTCTCTGCTCA 101
Db 190 GGCGTCACTGTGTCGAGAGGGGTGCGGATCTGTCTGTA 230

RESULT 11

US-09-105-537-1
; Sequence 1, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 15872
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-105-537-1

Query Match Best Local Similarity 72.9%; Score 54.4; DB 4; Length 15872;
Matches 70; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 3 CGCGCTGACAGGCTGCTCCGACGAGGCTGCAAGAGCTTCGCGCCGACGCGAG 62
Db 9136 CGCGCAGAGAGGCTGCGCCGACGCGCTGCAAGGCTTCTGCGCCGCGCGAG 9195
QY 63 CGTGGGTGAGAGGCTGCGCATCTCTCTGCT 98
Db 9196 CACCGGCTGGGCGGAGGCGTGGCTGTCTCT 9231

RESULT 12

US-08-804-227C-1
; Sequence 1, Application US/08804227C
; Patent No. 5876991
; GENERAL INFORMATION:
; APPLICANT: Dehoff, Bradley S.
; APPLICANT: Kunstoss, Stuart A.
; APPLICANT: Rostock, Paul R., Jr.
; APPLICANT: Sutton, Kimberly L.
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: THOMAS G. PLANT 1501
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCII(DOS) Text only
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804,227C
; FILING DATE: February 21, 1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Plant, Thomas, G.
; REGISTRATION NUMBER: 35,784
; REFERENCE/DOCKET NUMBER: X-8231
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-2459
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 43280 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)

FEATURE:
; NAME/KEY: CDS
; LOCATION: 816..14234
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 14351..19945
; FEATURE:
; NAME/KEY: CDS
; LOCATION: *20010..31199
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 31232..36067
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 36249..41774
US-08-804-227C-1

Query Match Best Local Similarity 72.9%; Score 54.4; DB 2; Length 43280;
Matches 70; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 3 CGCGCTGACAGGCTGCTCCGACGAGGCTGCAAGAGCTTCGCGCCGACGCGAG 62
Db 36998 CGCGCAGAGAGGCTGCGCCGACGCGCTGCAAGGCTTCTGCGCCGCGCGAG 37057
QY 63 CGTGGGTGAGAGGCTGCGCATCTCTCTGCT 98
Db 37058 CACCGGCTGGGCGGAGGCGTGGCTGTCTCT 37093

RESULT 13
US-09-036-987A-1

```

; Sequence 1, Application US/09036987A
; Patent No. 6143526
; GENERAL INFORMATION:
; APPLICANT: Baltz, Richard H.
; APPLICANT: Broughton, Mary C.
; APPLICANT: Crawford, Kathryn P.
; APPLICANT: Madduri, Krishnamurthy
; APPLICANT: Merlo, Donald J.
; APPLICANT: Treadway, Patti J.
; APPLICANT: Turner, Jan R.
; APPLICANT: Waldron, Clive
; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dow Agrosciences LLC Patent Department
; STREET: 9330 Zionsville Road
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: USA
; ZIP: 46268
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/036,987A
; FILING DATE: 09-Mar-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Stuart, Donald R.
; REGISTRATION NUMBER: 28,479
; REFERENCE/DOCKET NUMBER: 50,608
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (317)337-4816
; TELEFAX: (317)337-4847
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 80161 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-09-036-987A-1

Query Match 52.7%; Score 53.2; DB 3; Length 80161;
Best Local Similarity 71.4%; Pred. No. 2.5e-06;
Matches 70; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

QY 1 AGCGGCTGGAGGCTCCGAGCGGCAAGAGCTTCGCGCCGACGGCAG 60
Db 60538 AGCGGGAAGCGGCGCTCGGATGGCGGCAAGCGCTTCGGGATGCGCGCAG 60597
QY 61 GCGGTGGGTGAGAGCGTGGCGCATGCTCTGCT 98
Db 60598 GGCACCGCGTGGGCGGAGGCGCGCGAATGCTGCTGCT 60635

RESULT 14
; Sequence 1, Application US/09370700
; Patent No. 6274350
; GENERAL INFORMATION:
; APPLICANT: Baltz, Richard H.
; APPLICANT: Broughton, Mary C.
; APPLICANT: Crawford, Kathryn P.
; APPLICANT: Madduri, Krishnamurthy
; APPLICANT: Treadway, Patti J.
; APPLICANT: Turner, Jan R.
; APPLICANT: Waldron, Clive
; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
; FILE REFERENCE: 50489 Div1

```

```

; CURRENT APPLICATION NUMBER: US/09/370,700
; CURRENT FILING DATE: 1999-08-09
; EARLIER APPLICATION NUMBER: US 09/36987
; EARLIER FILING DATE: 1998-03-09
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 80161
; TYPE: DNA
; ORGANISM: Saccharopolyspora spinosa
; US-09-370-700-1

Query Match 52.7%; Score 53.2; DB 4; Length 80161;
Best Local Similarity 71.4%; Pred. No. 2.5e-06;
Matches 70; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

QY 1 AGCGGCTGGAGGCTCCGAGCGGCAAGAGCTTCGCGCCGACGGCAG 60
Db 60538 AGCGGGAAGCGGCGCTCGGATGGCGGCAAGCGCTTCGGGATGCGCGCAG 60597
QY 61 GCGGTGGGTGAGAGCGTGGCGCATGCTCTGCT 98
Db 60598 GGCACCGCGTGGGCGGAGGCGCGCGAATGCTGCTGCT 60635

RESULT 15
; Sequence 6, Application US/08258261B
; Patent No. 5639949
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James M.
; APPLICANT: Beck, James Joseph
; APPLICANT: Hill, Dwight Steven
; APPLICANT: Ryals, John Andrew
; APPLICANT: Gaffney, Thomas Deane
; APPLICANT: Lam, Stephen Ting
; APPLICANT: Hammer, Phillip E.
; APPLICANT: Uknes, Scott Joseph
; TITLE OF INVENTION: Genes for the synthesis of
; TITLE OF INVENTION: antipathogenic substances
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/258,261B
; FILING DATE: 08-JUN-1994
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/457,205
; FILING DATE: 01-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8614
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28958 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single

```

; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEITICAL: NO
; ANTI-SENSE: NO
US-08-258-261B-6

Query Match

52.1%; Score 52.6; DB 1; Length 28958;

Best Local Similarity 77.1%; Pred. No. 3.5e-06;

Matches 64; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 19 GCTCCGAGGAGGAGGTCGCAAGAGCTTCTGCGCCGACGCGCGGTGGGGTGAGCGAA 78

Db 15211 GCCCCGACGGTCGTCGCAAGGCCCTCTCGCCGGAAGCCGACGTCGGGCTGGCCGAA 15270

QY 79 GCGTGGCCATGCTCCTGCTCAA 101

Db 15271 GGGCCGGGATGCTCTGCTCGA 15293

Search completed: November 6, 2002, 16:05:45
Job time : 206.625 secs

RESULT 1	ID	AA55887	standard; DNA; 68750 BP.
XX	AA55887		
XX	AC		
XX	DT		
XX	10-APR-2000	(first entry)	
DE	Sorangium celulosum	68.75 kb contig.	
XX	Epithlone biosynthesis;	type I polyketide synthase;	taxol substitute;
KW	anticancer;	ds.	
XX			
OS	Sorangium celulosum.		
XX			
Key	Location/Qualifiers		
FT	1..1826		
FT	CDS		
FT	/partial		
FT	/product= "Partial Orf 1 protein (AAV58580)"		
FT	/note= "No initiation codon given in the specification"		
FT	complement (1900..3171)		
FT	/tag= b		
FT	/product= "Orf 2 protein (AAV58581)"		
FT	3415..5556		
FT	/tag= c		
FT	/product= "Orf 3 protein (AAV58582)"		
FT	complement (5612..5992)		
FT	/tag= d		
FT	/product= "Orf 4 protein (AAV58583)"		
FT	6226..6675		
FT	/tag= e		

```

FT      /product= "Orf 5 protein (AAV58584)"
FT      /tag= f
FT      /product= "Type I polyketide synthase, EPOS A
FT      11872..116104
FT      /tag= g
FT      /product= "Non-ribosomal peptide synthetase, EPOS P
FT      16251..21749
FT      /tag= h
FT      /product= "Type I polyketide synthase, EPOS B
FT      21746..43519
FT      /tag= i
FT      /product= "Type I polyketide synthase, EPOS C
FT      43524..54920
FT      /tag= j
FT      /product= "Type I polyketide synthase, EPOS D
FT      54935..62254
FT      /tag= k
FT      /product= "Type I polyketide synthase, EPOS E
FT      62369..63628
FT      /tag= l
FT      /product= "Cytochrome P450 oxygenase homologue, EPOS F
FT      63779..64333
FT      /tag= m
FT      /product= "Orf 6 protein (AAV58585)"
FT      complement (63853..64290)
FT      /tag= n
FT      /product= "Orf 7 protein (AAV58586)"
FT      64363..64920
FT      /tag= o
FT      /product= "Orf 8 protein (AAV58587)"
FT      complement (64287..64727)
FT      /tag= p
FT      /product= "Orf 9 protein (AAV58588)"
FT      65063..65767
FT      /tag= q
FT      /product= "Orf 10 protein"
FT      complement (65008..65874)
FT      /tag= r
FT      /product= "Orf 11 protein (AAV58590)"
FT      complement (65871..66338)
FT      /tag= s
FT      /product= "Orf 12 protein (AAV58591)"
FT      66667..67137
FT      /tag= t
FT      /product= "Orf 13 protein (AAV58592)"
FT      67334..68251
FT      /tag= u
FT      /product= "Orf 14 protein (AAV58593)"
FT      68346..68750
FT      /tag= v
FT      /product= "Partial Orf 15 protein (AAV58594)"
FT      /note= "No termination codon given in the specification"
XX      W09966028-A2.
XX      PD
XX      23-DEC-1999.
XX      PE
XX      16-JUN-1999; 99WO-EP04171.
XX      PR
XX      18-JUN-1998; 98US-0099504.
XX      PR
XX      24-SEP-1998; 98US-0101631.
XX      PR
XX      05-FEB-1999; 98US-0118906.
XX      PA
XX      (NOVS ) NOVARTIS AG.

```

```

PA      (NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.
XX      Schupp T, Ligon JM, Molnar I, Zirkle R, Goerlach J, Cyr D;
XX      WPI: 2000-097741/08.
XX      DR
XX      P-PSDB: AAV58573, AAV58574, AAV58575, AAV58576, AAV58577, AAV58578,
XX      AAV58579, AAV58580, AAV58581, AAV58582, AAV58583, AAV58584,
XX      AAV58585, AAV58586, AAV58587, AAV58588, AAV58590, AAV58591,
XX      AAV58592, AAV58593, AAV58594.
XX      PT
XX      New isolated epoethione synthase genes, used for the recombinant
XX      production of epoethione for use in cancer therapy.
XX      Claim 14; Page 87-104; 174pp; English.
XX      PS
XX      This sequence represents a 68.75 kb contig from Sorangium cellulosum
XX      comprising 22 open reading frames (ORFs) and includes genes encoding
XX      proteins involved in the biosynthesis of epoethiones. Epoethiones A and
XX      B are 16-membered macrocyclic polyketides with an acylcysteine-derived
XX      starter unit; polyketides being synthesised from two-carbon building
XX      blocks, the beta-carbon of which always carries a keto group. Each round
XX      of two-carbon addition is carried out by a complex of enzymes known as
XX      the polyketide synthase in a manner similar to fatty acid biosynthesis.
XX      EPOS A (AAV58573) and EPOS P (AAV58574) are involved in formation of
XX      the thiazole ring formation of epoethiones, and EPOS B, EPOS C, EPOS D
XX      and EPOS E (AAV58575-Y58578) are involved in polyketide backbone
XX      formation. EPO F (AAV58579) is an epoethione macrolactone oxidase, and
XX      the proteins Orf 3 (AAV58582) and Orf14 (AAV58593) are thought to be
XX      involved in transport. Epoethiones mimic the biological activity of
XX      taxol, and may be substituted for taxol in cancer chemotherapeutic
XX      compositions. Epoethiones exhibit a much lower drop in potency against a
XX      multiply drug-resistant cell line compared with taxol, and are
XX      considerably less efficiently exported from such cells by the multidrug
XX      resistance protein (MDR, or P-glycoprotein). Despite the potential of
XX      epoethiones as anticancer agents, they are problematical to produce on a
XX      large scale. Epoethiones are too complex for industrial scale chemical
XX      synthesis, and Sorangium cellulosum is difficult to ferment, producing
XX      poor yields of epoethiones. The nucleic acids of the invention may be
XX      used for the recombinant production of epoethiones in a heterologous host
XX      that is more amenable to fermentation.
XX      SQ
XX      Sequence 68750 BP; 9596 A; 22458 C; 25537 G; 11159 T; 0 other;
XX      Query Match
XX      Best Local Similarity 100.0%; Score 101; Length 68750;
XX      Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX      QY
XX      1 AGCCGCTCGAGAGCTGCTCCCGACGAGCGTGCAGAGCTTTCGCCGCCACCGAC 60
XX      |||||||
XX      DB 37682 AGCCGCTCGAGAGCTGCTCCCGACGAGCGTGCAGAGCTTTCGCCGCCACCGAC 3741
XX      QY
XX      61 GGGGTGGGGTGGAGCGAGCGTGGCCGATGCTCGCTCA 101
XX      |||||||
XX      DB 37742 GGGGTGGGGTGGAGCGAGCGTGGCCGATGCTCGCTCA 37782
XX      RESULT 2
XX      AAA29349
XX      ID AAA29349 standard; DNA; 71989 BP.
XX      AC
XX      AAA29349;
XX      DT
XX      12-SEP-2000 (first entry)
XX      DE
XX      Sorangium cellulosum epoethione polyketide synthase operon genomic DNA.
XX      KW
XX      Epoethione; polyketide synthase; epoa; epob; epoc; epod; epoe; epof;
XX      epog; epok; P450 epoxidase; ORFA; ORFB; promoter; enhancer; anti-fungal;
XX      tubulin polymerization assay; anti-tumour; cytostatic; ds.
XX      OS
XX      Sorangium cellulosum.
XX      FH
XX      Key Location/Qualifiers

```

```
FT CDS 3..992
FT /tag= a
FT /label= ORF_A
FT /product= transposase
FT /note= "not part of the PKS"
FT CDS 989..1501
FT /tag= b
FT /label= ORF_B
FT /product= transposase
FT /note= "not part of the PKS"
FT CDS 1998..6263
FT /tag= c
FT /label= epoa_gene
FT /note= "encodes the loading domain"
FT misc_RNA 2031..3548
FT /tag= d
FT /note= "encodes ketide synthase (KS-Y) of the loading
FT domain"
FT misc_RNA 3621..4661
FT /tag= e
FT /note= "encodes acyl transferase (AT) of the loading
FT domain"
FT misc_RNA 4917..5810
FT /tag= f
FT /note= "encodes enoyl reductase (ER) of the loading
FT domain, potentially involved in formation of the
FT thiazole moiety"
FT misc_RNA 5856..6155
FT /tag= g
FT /note= "encodes acyl carrier protein (ACP) of the loading
FT domain"
FT CDS 6260..10493
FT /tag= h
FT /label= epob_gene
FT /note= "encodes module 1, the NRPS module"
FT misc_RNA 2031..3548
FT /tag= i
FT /note= "encodes condensation domain C2 of the NRPS
FT module"
FT misc_RNA 2031..3548
FT /tag= j
FT /note= "encodes condensation domain C2 of the NRPS
FT module"
FT misc_RNA 6861..6887
FT /tag= k
FT /note= "encodes heterocyclization signature sequence"
FT misc_RNA 6861..6887
FT /tag= l
FT /note= "encodes condensation domain C4 of the NRPS
FT module"
FT misc_RNA 7358..7366
FT /tag= m
FT /note= "encodes condensation domain C7 (partial) of the
FT NRPS module"
FT misc_RNA 7898..7921
FT /tag= n
FT /note= "encodes adenylation domain A1 of the NRPS module"
FT misc_RNA 7898..7921
FT /tag= o
FT /note= "encodes adenylation domain A1 of the NRPS module"
FT misc_RNA 8261..8308
FT /tag= p
FT /note= "encodes adenylation domain A3 of the NRPS module"
FT misc_RNA 8411..8422
FT /tag= q
FT /note= "encodes adenylation domain A4 of the NRPS module"
FT misc_RNA 8861..8905
FT /tag= r
FT /note= "encodes adenylation domain A6 of the NRPS module"
FT misc_RNA 8966..8983
FT /tag= s
FT /note= "encodes adenylation domain A7 of the NRPS module"
FT misc_RNA 9090..9179
FT /tag= t
FT /note= "encodes adenylation domain A8 of the NRPS module"
FT misc_RNA 9183..9992
FT /tag= u
FT /note= "encodes oxidation region for forming thiazole"
FT misc_RNA 10121..10138
FT /tag= v
FT /note= "encodes adenylation domain A10 of the NRPS
FT module"
FT misc_RNA 10261..10306
FT /tag= w
FT /note= "encodes thiolation domain (PCP) of the NRPS
FT module"
FT CDS 10639..16137
FT /tag= x
FT /label= epoc_gene
FT /note= "encodes module 2"
FT misc_RNA 10654..12033
FT /tag= y
FT /note= "encodes KS2, the KS domain of module 2"
FT misc_RNA 12250..13287
FT /tag= z
FT /note= "encodes AT2, the AT domain of module 2"
FT misc_RNA 13327..13899
FT /tag= aa
FT /note= "encodes dehydratase (DH) 2, the DH domain of
FT module 2"
FT misc_RNA 14962..15756
FT /tag= ab
FT /note= "encodes ketoreductase (KR) 2, the KR domain of
FT module 2"
FT misc_RNA 15763..16008
FT /tag= ac
FT /note= "encodes ACP2, the ACP domain of module 2"
FT CDS 16134..37907
FT /tag= ad
FT /label= epod_gene
FT /note= "encodes modules 3-6"
FT misc_RNA 16425..17606
FT /tag= ae
FT /note= "encodes KS3"
FT misc_RNA 17817..18857
FT /tag= af
FT /note= "encodes AT3"
FT misc_RNA 19581..20396
FT /tag= ag
FT /note= "encodes KR3"
FT misc_RNA 20424..20642
FT /tag= ah
FT /note= "encodes ACP3"
FT misc_RNA 20706..22082
FT /tag= ai
FT /note= "encodes KS4"
FT misc_RNA 22296..23336
FT /tag= aj
FT /note= "encodes AT4"
FT misc_RNA 24059..24647
FT /tag= ak
FT /note= "encodes KR4"
FT misc_RNA 24867..25151
FT /tag= al
FT /note= "encodes ACP4"
FT misc_RNA 25203..26576
FT /tag= am
FT /note= "encodes KS5"
FT misc_RNA 26793..27883
FT /tag= an
FT /note= "encodes AT5"
FT misc_RNA 27966..28574
FT /tag= ao
FT /note= "encodes DH5"
FT misc_RNA 29433..30287
FT /tag= ap
```

FT	/note=	"encodes ER5"
FT	30321..	30869
FT	misc_RNA	
FT	/*tag=	aq
FT	/note=	"encodes KR5"
FT	31077..	31373
FT	misc_RNA	
FT	/*tag=	ar
FT	/note=	"encodes ACP5"
FT	31440..	32807
FT	misc_RNA	
FT	/*tag=	as
FT	/note=	"encodes KS6"
FT	33018..	34067
FT	misc_RNA	
FT	/*tag=	at
FT	/note=	"encodes AT6"
FT	34107..	34676
FT	misc_RNA	
FT	/*tag=	au
FT	/note=	"encodes DH6"
FT	35760..	36641
FT	misc_RNA	
FT	/*tag=	av
FT	/note=	"encodes ER6"
FT	36705..	37256
FT	misc_RNA	
FT	/*tag=	aw
FT	/note=	"encodes KR6"
FT	37470..	37769
FT	misc_RNA	
FT	/*tag=	ax
FT	/note=	"encodes ACP6"
FT	37912..	49308
FT	CDS	
FT	/*tag=	ay
FT	/label=	epoE_gene
FT	/note=	"encodes modules 7 and 8"
FT	38014..	39375
FT	misc_RNA	
FT	/*tag=	az
FT	/note=	"encodes KS7"
FT	39589..	40626
FT	misc_RNA	
FT	/*tag=	ba
FT	/note=	"encodes AT7"
FT	41341..	41922
FT	misc_RNA	
FT	/*tag=	bb
FT	/note=	"encodes KR7"
FT	42181..	42423
FT	misc_RNA	

Query Match	100.0%	Score 101:	DB 21:	Length 71989:
Best Local Similarity	100.0%	Pred. No. 1.3e-18:		
Matches 101:	Conservative 0:	Mismatches 0:	Indels 0:	Gaps 0
Qy 1	AGCGCGCTGCCAGAGCCTGAGCTCCGACGACGGTGCACAGAGCTTCTCGGCCGACGCCGAC	60		
Db 32070	AGCGCGCTGCCAGAGCCTGAGCTCCGACGACGGTGCACAGAGCTTCTCGGCCGACGCCGAC	32129		
Qy 61	GCGCTGGGGTGGAGGAGCGAGCTGCCCATGCTTCCTCCGCTCAA	101		
Db 32130	GCGCTGGGGTGGAGGAGCGAGCTGCCCATGCTTCCTCCGCTCAA	32170		
RESULT 3				
ACCS5785				
ID	ACCS5785 standard; DNA: 3978 BP.			
AC				
ACCS5785:				
AC				
DT	19-JAN-2001 (first entry)			
XX				
DE	Type I polyketide synthase orf8.			
XX				
KW	Mitomycin; biosynthesis; mitosome ring system; antibiotic; anti-cancer;			
KW	anti-inflammatory; immune-enhancer; immunosuppressant; asthma;			
KW	chronic obstructive pulmonary disease; respiratory inflammation;			
XX	fungicide; pesticide; ds.			
XX				
OS	Streptomyces lavendulae.			
XX				
PN	MO200053737-A2.			
XX				
DD	14-SEP-2000.			

XX 10-MAR-2000; 2000MO-US06394.
 PF
 XX
 PR 12-MAR-1999; 99US-0266965.
 XX
 PA (MINU) UNIV MINNESOTA.
 PA (SHER/) SHERMAN D H.
 PA (MAOY/) MAO Y.
 PA (VARO/) VAROGLU M.
 PA (HEMW/) HE M.
 PA (SHEL/) SHELDON P C.
 XX
 PI Sherman DH, Mao Y, Varoglu M, He M, Sheldon PC;
 XX
 DR WPI: 2000-601980/57.
 XX
 PT Novel nucleic acid molecule comprising mitomycin biosynthetic gene
 PT cluster useful for cloning mitomycin biosynthetic genes for elucidating
 PT the molecular basis of mitosome ring system biosynthesis
 XX
 PS Example 1; Page 251-253; 399pp; English.

CC This invention relates to isolated and purified nucleic acid molecules
CC from the mitomycin biosynthetic gene cluster. Mitomycins are a group of
CC natural products that contain a variety of functional groups, including
CC amino benzozulone and axiridine ring systems. The *S. lavendulae*
CC mitomycin biosynthetic gene cluster comprises 47 mitomycin genes
CC spanning 55 kb of DNA. The invention includes an expression cassette
CC comprising a mitomycin biosynthetic gene operably linked to a promoter,
CC and host cells transformed with the cassette. The nucleotide, and protein
CC sequences and the transformed host cells of the invention result in
CC antitastmatic, antiinflammatory, cytostatic, immunomodulatory, and
CC antibiotic activities. The nucleotide sequences are used to elucidate the
CC molecular basis for the biosynthesis of the mitosome ring system, as well
CC as to engineer the biosynthesis of novel natural products, e.g.,
CC antibiotics, anti-inflammatory agents, anti-cancer agents,
CC immune-enhancers, immunosuppressants, agents to treat asthma, chronic
CC obstructive pulmonary disease as well as other disease involving
CC respiratory inflammation, or cholesterol-lowering agents or as crop
CC protection agents (e.g. fungicides or insecticides) as well as
CC biopolymers, e.g., in packaging or biomedical applications, or to engineer
CC PAA monomer syntheses. Sequences AAC55782-C55881, AAC55815-C55849 and
CC ABAB32485-B35542 represent mitomycin biosynthetic gene clusters DNA
CC sequences and encoded proteins. Sequences AAC55812-C55814,
CC AAC55850-C55856 and AAC55862-C55869 represent PCR primers used in the
CC cloning of the mitomycin biosynthetic genes.

```

Query Match      57.4% Score 58: DB 21: Length 3978:
Best Local Similarity 74.5% Pred No. 6,1e-07:
Matches 73: Conservative 0: Mismatches 25: Indels 0: Gaps 0

QY      1  AGCCGCGCTGCAGAGCCTGCGCTCCCGACGGAGCGTCGACAGAGCTTCTCGCGCCGACGCCGAC 60
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db       742  AGCCGCGCAGGGGGGCTGGGCCCGCCGACGGCGCGGTGACATCTCTTCGCGCACGCCCGCAC 801

QY      61  GCGCTGGGGGTGAGCCGAAAGGCTCGCCCATGCTCTGCT 98
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db       802  GGCACGACCTGGTCCGAGGGCGGGCGGCTGCTGCT 839

RESULT 4
AAC55840
ID    AAC55840 standard; DNM; 12249 BP.
XX
XX    AAC55840:
XX
XX    19-JAN-2001 (first entry)
XX
XX    Complete Mitomycin ORF 1-9 nucleotide sequence.
XX
XX    Mitomycin: biosynthesis; mtosomal ring system; antibiotic; anti-cancer;

```

KW		anti-inflammatory; immune-enhancer; immunosuppressant; asthma;
KM		chronic obstructive pulmonary disease; respiratory inflammation;
RN		fungicide; pesticide; ds.
XX		
OS		Streptomyces lavendulae.
PN		WO200053737-A2.
PD		
XX		14-SEP-2000.
PF		
XX		10-MAR-2000; 2000WO-US06394.
PR		
XX		12-MAR-1999; 99US-0266965.
PA		(MINU) UNIV MINNESOTA.
XX		(SHER/) SHERMAN D H.
PA		(MAOY/) MAO Y.
PA		(VARO/) VAROGIU M.
PA		(HEMM/) HE M.
XX		(SHEL/) SHELDON P C.
PI		Sherman DH, Mao Y, Varogiu M, He M, Sheldon PC;
DR		
XX		WPI: 2000-601980/57.
PT		
XX		
PS		Claim 26; Figure 22; 399pp; English.
XX		
CC		This invention relates to isolated and purified nucleic acid molecules
CC		from the mitomycin biosynthetic gene cluster. Mitomycins are a group of
CC		natural products that contain a variety of functional groups, including
CC		amino benzocounone and axitidine ring systems. The S. lavendulae
CC		mitomycin biosynthetic gene cluster comprises 47 mitomycin genes
CC		spanning 55kb of DNA. The invention includes an expression cassette
CC		comprising a mitomycin biosynthetic gene operably linked to a promoter,
CC		and host cells transformed with the cassette. The nucleotide, and protein
CC		sequences and the transformed host cells of the invention result in
CC		antipneumonia, antiinflammatory, cytostatic, immunomodulatory, and
CC		antibiotic activities. The nucleotide sequences are used to elucidate the
CC		molecular basis for the biosynthesis of the mitosome ring system, as well
CC		as to engineer the biosynthesis of novel natural products, e.g.
CC		antibiotics, anti-inflammatory agents, anti-cancer agents,
CC		immune-enhancers, immunosuppressants, agents to treat asthma, chronic
CC		obstructive pulmonary disease as well as other disease involving
CC		respiratory inflammation, or cholesterol-lowering agents or as crop
CC		protection agents (e.g. fungicides or insecticides) as well as
CC		dipolymer, e.g., in packaging or biomedical applications, or to engineer
CC		PMA monomer synthases. Sequences AAC55787-C55881, AAC55815-C55849 and
CC		AAC32485-332542 represent mitomycin biosynthetic gene cluster DNA
CC		sequences and encoded proteins. Sequences AAC55812-C55814,
CC		AAC55850-C55856 and AAC55862-C55869 represent PCR primers used in the
CC		cloning of the mitomycin biosynthetic genes.
XX		
SO		Sequence 12249 BP; 1715 A; 4722 C; 4196 G; 1616 T; 0 other;
NY		
Query Match	57.4%;	Score 58; DB 21; Length 12249;
Best Local Similarity	74.5%;	Pred. No. 6.4e-07;
Matches 73; Conservative	0;	Mismatches 25; Indels 0; Gaps 0
OY	1	AGCGGCTGCAGAGCCTGGCTCCCAAGACAAGTGTGAAGAAGCTTCGCGCGACCGGAC 60
DB	1729	AGCGGCAAGCGGGGCTGGCCCCCGACAGGCGCCTTGATCCTTCGCGGACGCCGAC 1788
OY	61	GCGTGGGTGAGCGAAGCTGCGCATGCTCTGCT 98
DB	1789	GCGCACCTGTGTCGAGGCGCGCGCGTGTGCTCT 1826

ID	ACCS5857	standard; DNA; 18331 BP.
XX		
AC	AACS5857;	
XX		
DT	19-JAN-2001	(first entry)
XX		
DE	Complete nucleotide sequence of the mitomycin gene cluster.	
XX		
XX	Mitomycin; biosynthesis; mitosome ring system; antibiotic; anti-cancer;	
KW	anti-inflammatory; immune-enhancer; immunosuppressant; asthma;	
KW	chronic obstructive pulmonary disease; respiratory inflammation;	
KW	fungicide; pesticide; ds.	
XX		
OS	Streptomyces lavendulae.	
XX		
PN	WO20053737-A2.	
XX		
PD	14-SEP-2000.	
XX		
PF	10-MAR-2000; 2000WO-US06394.	
XX		
PR	12-MAR-1999; 99US-0266965.	
XX		
PA	(MNU) UNIV MINNESOTA.	
PA	(SHER/) SHERMAN D H.	
PA	(MAOY/) MAO Y.	
PA	(VARO/) VAROGLU M.	
PA	(HEM/) HE M.	
PA	(SHEL/) SHELTON P C.	
XX		
PI	Sherman DH, Mao Y, Varoglu M, He M, Sheldon PC;	
XX		
DR	WPI, 2000-601980/57.	
XX		
PT	Novel nucleic acid molecule comprising mitomycin biosynthetic gene	
XX	cluster useful for cloning mitomycin biosynthetic genes for elucidating	
PT	the molecular basis of mitosome ring system biosynthesis	
XX		
PS	Disclosure: Figure 21; 399pp; English.	
XX		
CC	This invention relates to isolated and purified nucleic acid molecules	
CC	from the mitomycin biosynthetic gene cluster. Mitomycins are a group of	
CC	natural products that contain a variety of functional groups, including	
CC	amino benzoquinone and axiridine ring systems. The S. lavendulae	
CC	mitomycin biosynthetic gene cluster comprises 47 mitomycin genes	
CC	spanning 55kb of DNA. The invention includes an expression cassette	
CC	comprising a mitomycin biosynthetic gene operably linked to a promoter,	
CC	and host cells transformed with the cassette. The nucleotide, and protein	
CC	sequences and the transformed host cells of the invention result in	
CC	antitlastic, antiinflammatory, cytostatic, immunomodulatory, and	
CC	antibiotic activities. The nucleotide sequences are used to elucidate the	
CC	molecular basis for the biosynthesis of the mitosome ring system, as well	
CC	as to engineer the biosynthesis of novel natural products, e.g.	
CC	antibiotics, anti-inflammatory agents, anti-cancer agents,	
CC	immune-enhancers, immunosuppressants, agents to treat asthma, chronic	
CC	obstructive pulmonary disease as well as other disease involving	
CC	respiratory inflammation, or cholesterol-lowering agents or as crop	
CC	protection agents (e.g. fungicides or insecticides) as well as	
CC	biopolymers, e.g., in packaging or biomedical applications, or to engineer	
CC	PHA monomer synthases. Sequences AAC5782-C55881, AAC55815-C55849 and	
CC	AAB32445-B32542 represent mitomycin biosynthetic gene cluster DNA	
CC	sequences and encoded proteins. Sequences AAC55812-C55814,	
CC	AAC55850-C55856 and AAC55862-C55869 represent PCR primers used in the	
CC	cloning of the mitomycin biosynthetic genes.	
XX		
SO	Sequence 18331 BP; 2523 A; 7003 C; 6343 G; 2462 T; 0 other;	
XX		
Query Match	57.4%;	Score 58; DB 21; Length 18331;
Best Local Similarity	74.5%;	Pred. No. 6-5e-07;
Matches	73; conservative	0; Mismatches 25; Indels 0; Gaps 0;

1 AGCCGAGTCGAGGCGCTGCGCCGACGAGCGATGCGACAGAGCTTCCTGGCCGCGACGCCGAC 60
||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 1729 AGCCGACGAGCGGGGCTGGCCCCCGACGGCCGTCGACAGTCTTTCGGCCGACGCCGAC 1788
Qy 61 GGCGTGGGTGAGAGGAGCTGCGCCATGCTCTGCT 98
111 111 111 111 111 111
Db 1789 GGCACCACTGCTCCAGGGCGGCGCTCTCTCT 1826

RESULT 6
AAF24892
ID AAF24892 standard; DNA; 20394 BP.

XX AAF24892;

XX 20-APR-2001 (first entry)

DE Pimaricin biosynthesis associated polyketide synthase gene.

KW Polyketide synthase; oxidative modification; metabolite; antibiotic;

KM anticancer; pimaricin; ss.

OS Streptomyces natalensis.

XX Key Location/Qualifiers

FT CDS 1..20394

FT /tag= a

FT /product= "polyketide synthase"

PN MO200077222-A1.

PD 21-DEC-2000.

PF 14-JUN-2000; 2000WO-EP06227.

PR 14-JUN-1999; 99EP-0201893.

PA (STM) DSM NV.

PI Martin JF, Aparicio JF, Colina AJ;

DR MPI: 2001-080693/09.

XX P-PSDB: AAB31558.

PT New polynucleotides encoding enzymes involved in the biosynthesis of
PM pimaricin, useful for modifying the biosynthesis of pimaricin and in
the synthesis of new compounds -

PS Disclosure: Page 53-80; 116pp; English.

CC The present sequence encodes a polyketide synthase which is associated
CC with the biosynthesis of pimaricin. The polyketide synthase polypeptide
CC is useful for the oxidative modification of a methyl group of a suitable
CC compound, e.g. a bioactive compound including a secondary metabolite,
CC antibiotics and anticancer agents. Recombinant cells comprising the
CC gene are useful for the production of pimaricin. The polyketide synthase
CC polynucleotide may be over expressed in Streptomyces, leading to an
CC increase in the biosynthesis of pimaricin, as a source of primers for
CC amplification reaction and as probes.

XX Sequence 20394 BP; 2611 A; 7486 C; 7397 G; 2900 T; 0 other;

Query Match 57.2%; Score 57.8; DB 22; Length 20394;
Best Local Similarity 76.3%; Pred. No. 7.4e-07;
Matches 71; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

Qy 3 CGGGCTGCAGAGGCTGCTCCGACGAGGTCAGAGACTTCTGCGCCGACCGACG 62
1111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
Db 5319 CGGGACGCGCGGCTGGCCCGACGCGCTTCGCGCGCGCCGCGACG 5378

Qy 63 CGTGGGTGAGAGGCTGCGCCATGCTCT 95
1 11 111111 111 11 1111 11
Db 5379 CACCGGCTGAGAGGCGGCTGCTCT 5411

RESULT 7
AAF89964
ID AAF89964 standard; DNA; 672 BP.

XX AAF89964;

XX 06-AUG-2001 (first entry)

DE Partial nucleotide sequence of a type I polyketide synthase.

KW Metabolic pathway operon; polyketide; polyketide antibiotic;

KM type I polyketide synthase; ss.

XX Streptomyces ambofaciens.

OS MO2001A0497-A2.

PN 07-JUN-2001.

PD 27-NOV-2000; 2000WO-FR03311.

PF 29-NOV-1999; 99EP-0015032.

PR 07-JUN-2000; 2000US-0203800.

PA (AVER) AVENTIS PHARMA SA.

PI Jeannin P, Pernodet J, Guerneau M, Simonet P, Courtols S;

DR Cappellano C, Francou F, Raynal A, Ball M, Sezonov G, Tuphile K;

PI Frostegard A;

DR MPI: 2001-374849/39.

XX P-PSDB: AAB83956.

PT Collection of nucleic acids from environmental samples, useful for
identifying e.g. genes encoding polyketide synthases and derived
antibiotics -

PS Claim 35; Page 219-220; 356pp; French.

CC The specification describes a method for the preparation of a collection
CC of nucleic acids from organisms in a soil sample. The method comprises
CC milling a dried sample to produce microparticles; suspending these in
CC liquid buffer; extraction of nucleic acids from the microparticle;
CC passing nucleic acid-containing solution through a molecular sieve;
CC passing nucleic acid-enriched fractions through an anion exchange
CC chromatography material; and recovering fractions containing purified
CC nucleic acids. The nucleic acids are sources for sequences that encode
CC either operons involved in a metabolic pathway (specifically polyketide
CC synthesis) or polypeptides, particularly for production of therapeutic
CC or agricultural compounds, especially polyketide antibiotics. AAF89964-78
CC encode partial type I polyketide synthases, and were isolated using the
CC method of the invention.

XX Sequence 672 BP; 84 A; 238 C; 260 G; 90 T; 0 other;

Query Match 57.0%; Score 57.6; DB 22; Length 672;
Best Local Similarity 75.0%; Pred. No. 7.3e-07;
Matches 72; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

Qy 3 CGGGCTGCAGAGGCTGCTCCGACGAGGTCAGAGACTTCTGCGCCGACCGACG 62
1111 1 111111 1111 1 1111 111111 111111 111111
Db 378 CGGGACGCGCGGCTGGCCCGACGAGCGCGCTTCGCGCGCGACCG 437

Qy 63 CGTGGGTGAGAGGCTGCGCCATGCTCT 98
1 1111 1111 111 111 111 111 111 111 111 111 111 111 111
Db 438 CACGGGCTGGGCGAGGCGCTGCGGCTGCTCT 473

RESULT 8
AAS17367
ID AAS17367 standard; DNA; 33529 BP.
XX AAS17367;

```
XX 12-MAR-2002 (first entry)
XX
XX
DE DNA sequence of S. cellulosum polyketide synthase cosmid, PKOS28-26.
XX
XX Polyketide synthase; PKS; catalytic domain; ketosynthase domain;
XX acyl transferase domain; dehydratase domain; ketoreductase domain;
XX acyl carrier protein domain; PKOS28-26; ds.
XX
XX Sorangium cellulosum.
XX
XX US6280999-B1.
XX
XX 28-AUG-2001.
XX
XX 31-AUG-1998; 98US-0144085.
XX
XX 22-JAN-1998; 98US-0010809.
XX
XX (KOSA-) KOSAN BIOSCIENCE.
XX
XX Gustafsson C, Belach MC, Ashley G, Julien B, Ziemann R;
XX
XX WPI; 2001-606536/69.
XX
XX
XX Novel purified, isolated DNA molecule from Sorangium cellulosum having
XX polyketide open reading frame encoding modules with one or more domains
XX such as ketosynthase, acyl transferase and acyl carrier protein domains
XX
XX
XX Claim 4; Fig 1; 72pp; English.
XX
XX
XX The present invention relates to the isolation of novel Sorangium
XX cellulosum polyketide synthases (PKS), and the polynucleotide sequences
XX encoding them. The polyketide synthases include catalytic domains such
XX as ketosynthase domain, acyl transferase domain, dehydratase domain,
XX ketoreductase domain and acyl carrier protein domain. A host cell
XX comprising a PKS ORF (open reading frame) which encodes one or more
XX more PKS domains is useful for producing polyketide synthases from which
XX polyketides can be produced. The host cells are useful for constructing
XX a library, where each individual colony of the library represents a
XX colony with the ability to produce a particular PKS synthase and
XX ultimately a particular polyketide. The polyketides produced by these
XX colonies can be used collectively in a panel to represent a library or
XX may be assessed individually for activity. Colonies in the library are
XX also induced to produce the relevant synthases and thus to produce the
XX relevant polyketides to obtain a library of candidate polyketides which
XX can be screened for binding to desired targets such as receptors,
XX signalling proteins, etc. The present sequence represents the DNA
XX sequence of cosmid PKOS28-26 which encodes one or more domains of
XX S. cellulosum PKS.
XX
XX Note: The present sequence is said to encode the functional domains
XX of S. cellulosum PKS which correspond to domains or domain subsets of
XX the amino acid sequences of ORF1 (AAU10700) and ORF2 (AAU10701).
XX
XX
XX Sequence 33529 BP; 4489 A; 9518 C; 14470 G; 5046 T; 6 other:
XX
XX
XX Query Match 55.68; Score 56.2; DB 23; Length 33529;
XX Best Local Similarity 72.38; Pred. No. 2.1e-06;
XX Matches 73; Conservative 0; Mismatches 28; Indels 0; Gaps 0;
XX
XX
XX 1 AGCCGCTGCGAGCTGCTCCGAGGAGCGTGCAGACCTCTCGGCCGACGCGAC; 60
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX Db 376 AGCCGCTCAAGGGGATGCGCCGCGACGCGCTGCAAGACCTCTCTGCGCGAGCTGAC; 435
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX Qy 61 GCGCTGGGCTGACGAGGCTGCGCATCTCTCTGCTCAA 101
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX Db 436 GCGCTCACCTGTCTCGAGGGGTGCGGATGCTGTGGTCTGAA 476
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX
XX
XX RESULT 9
XX AAF30757
XX ID AAF30757 standard; DNA; 47981 BP.
XX
XX
XX
XX AAF30757;
XX
XX 21-JUN-2001 (first entry)
XX
XX Micromonospora megalomicin biosynthetic gene cluster.
XX
XX Megalomicin; meg gene; polyketide synthase; antibiotic;
XX motillide; antiparasitic; ds.
XX
XX Micromonospora megalomicia subsp. nigra.
XX
XX Key
XX CDS
XX
XX Location/Qualifiers
XX complement (1..144)
XX /tag= a
XX /partial
XX /gene= "megBVI(megI)"
XX /product= "TDP-4-keto-6-deoxyglucose-
XX 2,3-dehydratase"
XX /note= "encodes AAB82201"
XX 928..2061
XX /tag= b
XX /gene= "megDVI"
XX /product= "TDP-4-keto-6-deoxyhexose 3,4-isomerase"
XX /note= "encodes AAB82202"
XX 2072..3382
XX /tag= c
XX /gene= "megDI"
XX /product= "TDP-megosamine glycosyltransferase"
XX /note= "eryCIII homologue; encodes AAB802203"
XX 3462..4634
XX /tag= d
XX /gene= "megI"
XX /product= "mycarose O-acyltransferase"
XX /note= "encodes AAB82204"
XX 4651..5775
XX /tag= e
XX /gene= "megDII"
XX /product= "TDP-3-keto-6-deoxyhexose
XX 3-aminotransaminase"
XX /note= "eryCI, DnrI homologue; encodes AAB82205"
XX 5822..6595
XX /tag= f
XX /gene= "megDIII"
XX /product= "daunosaminyl-N,N-dimethyltransferase"
XX /note= "eryCVI homologue; encodes AAB82206"
XX 6592..7197
XX /tag= g
XX /gene= "megDIV"
XX /product= "TDP-4-keto-6-deoxyhexose 3,5-epimerase"
XX /note= "eryBVIII, dnmU homologue; encodes AAB82207"
XX 7220..8206
XX /tag= h
XX /gene= "megDV"
XX /product= "TDP-4-keto-6-deoxyhexose 4-ketoreductase"
XX /note= "eryVIV, dnmV homologue; encodes AAB82208"
XX complement (8228..9220)
XX /tag= i
XX /gene= "megBII-1(megDVII)"
XX /product= "TDP-4-keto-6-deoxyhexose 2,3-reductase"
XX /note= "encodes AAB82209"
XX complement (9226..10479)
XX /tag= j
XX /gene= "megBV"
XX /product= "TDP-mycarose glycosyltransferase"
XX /note= "encodes AAB82210"
XX complement (10483..11424)
XX /tag= k
XX /gene= "megBIV"
XX /product= "TDP-4-keto-6-deoxyhexose 4-ketoreductase"
XX /note= "encodes AAB82211"
XX 12181..22821
XX /gene= "mega"
```

```

FT      /tag= l
FT      /product= "megalomycin 6-deoxyerythronolide B
FT      /note= "polyketide synthase; encodes AAB82212"
FT      /tag= m
FT      /gene= "mega"
FT      /function= "AT-L"
FT      /tag= n
FT      /gene= "mega"
FT      /function= "ACP-L"
FT      /tag= o
FT      /gene= "mega"
FT      /function= "KSI"
FT      /tag= p
FT      /gene= "mega"
FT      /function= "AT1"
FT      /tag= q
FT      /gene= "mega"
FT      /function= "KRI"
FT      /tag= r
FT      /gene= "mega"
FT      /function= "ACP1"
FT      /tag= s
FT      /gene= "mega"
FT      /function= "KS2"
FT      /tag= t
FT      /gene= "mega"
FT      /function= "AT2"
FT      /tag= u
FT      /gene= "mega"
FT      /function= "KR2"
FT      /tag= v
FT      /gene= "mega"
FT      /function= "ACP2"
FT      /tag= w
FT      /gene= "mega"
FT      /product= "megalomycin 6-deoxyerythronolide B
FT      /note= "polyketide synthase, encodes AAB82213"
FT      /tag= x
FT      /gene= "mega"
FT      /function= "KS3"
FT      /tag= y
FT      /gene= "mega"
FT      /function= "AT3"
FT      /tag= z
FT      /gene= "mega"
FT      /function= "KR3 (inactive)"
FT      /tag= aa
FT      /gene= "mega"
FT      /function= "ACP3"
FT      /tag= ab
FT      /gene= "mega"
FT      /function= "KS4"
FT      /tag= ac
FT      /gene= "mega"
FT      /tag= ad
FT      /gene= "mega"
FT      /function= "AT4"
FT      /tag= ae
FT      /gene= "mega"
FT      /function= "ER4"
FT      /tag= af
FT      /gene= "mega"
FT      /function= "KR4"
FT      /tag= ag
FT      /gene= "mega"
FT      /function= "ACP4"
FT      /tag= ah
FT      /gene= "mega"
FT      /product= "megalomycin 6-deoxyerythronolide B
FT      /note= "polyketide synthase; encodes AAB82214"
FT      /tag= ai
FT      /gene= "mega"
FT      /function= "KS5"
FT      /tag= aj
FT      /gene= "mega"
FT      /function= "AT5"
FT      /tag= ak
FT      /gene= "mega"
FT      /function= "KR5"
FT      /tag= al
FT      /gene= "mega"
FT      /function= "ACP5"
FT      /tag= am
FT      /gene= "mega"
FT      /function= "KS6"
FT      /tag= an
FT      /gene= "mega"
FT      /function= "AT6"
FT      /tag= ao
FT      /gene= "mega"
FT      /function= "KR6"
FT      /tag= ap
FT      /gene= "mega"

```

```

FT      /function= "AT4"
FT      /tag= ad
FT      /gene= "mega"
FT      /function= "DH4"
FT      /tag= ae
FT      /gene= "mega"
FT      /function= "ER4"
FT      /tag= af
FT      /gene= "mega"
FT      /function= "KR4"
FT      /tag= ag
FT      /gene= "mega"
FT      /function= "ACP4"
FT      /tag= ah
FT      /gene= "mega"
FT      /product= "megalomycin 6-deoxyerythronolide B
FT      /note= "polyketide synthase; encodes AAB82214"
FT      /tag= ai
FT      /gene= "mega"
FT      /function= "KS5"
FT      /tag= aj
FT      /gene= "mega"
FT      /function= "AT5"
FT      /tag= ak
FT      /gene= "mega"
FT      /function= "KR5"
FT      /tag= al
FT      /gene= "mega"
FT      /function= "ACP5"
FT      /tag= am
FT      /gene= "mega"
FT      /function= "KS6"
FT      /tag= an
FT      /gene= "mega"
FT      /function= "AT6"
FT      /tag= ao
FT      /gene= "mega"
FT      /function= "KR6"
FT      /tag= ap
FT      /gene= "mega"

```

Query Match 54.3%; Score 54.8; DB 22; Length 47981;
 Best Local Similarity 72.4%; Pred. No. 56-06;
 Matches 71; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

```

QY      1 AGCCGCTCGAGGCTGCTCCGACGAGCGTGCAAGAGCTTCGCGCCGACGCGAC 60
DB      14485 AGTCGGATGAACTCCTCCGCCCGACGAGCGTCCAGCGCTTCGCGCCGCGAC 14544
QY      61 GCGCTGGGTGAGCGAGCGATGCGGCTGCTGCT 98
DB      14545 GGGTTCGCATGGCCGAGGCGAGGATGCTCTCT 14582

```

RESULT 10
 ID AAA58914 standard; DNA; 751 BP.
 XX AAA58914;
 AC
 XX


```
DT 20-OCT-2000 (first entry)
XX
DE DNA encoding ketosynthase domain of epothilone polyketide synthase.
XX
KW Epothilone polyketide synthase; PKS; ketosynthase; acyltransferase;
KW acyl carrier protein; polyketide; ketoreductase; enoylreductase;
KW dehydratase; epothilone; antibiotic; antitumor; ss.
XX
OS Sorangium cellulosum.
XX
FH Key Location/Qualifiers
FT CDS 1..751
FT /tag= a
FT /transl_except= (pos: 124..126, aa: Xaa)
FT /note= "Xaa is any amino acid"
XX
XX US6090601-A.
XX
XX 18-JUL-2000.
XX
XX 23-JAN-1998; 98US-0010809.
XX
XX 23-JAN-1998; 98US-0010809.
XX
XX (KOSA-) KOSAN BIOSCIENCE.
XX
XX Gustafsson C, Betlach MC;
XX
XX WPI; 2000-498261/44.
XX
XX P-PSDB; AAB07517.
XX
XX Novel epothilone polyketide synthase comprising a ketoreductase or
XX ketosynthase domain, useful for producing epothilone -
XX
XX Claim 6; Column 15-16; 39pp; English.
XX
XX The present sequence encodes a ketosynthase domain of epothilone
XX polyketide synthase (PKS). In type I or modular PKS enzymes, a set of
XX separate catalytic active sites (each termed a domain. A set of which
XX is termed a module) exists for each cycle of carbon chain elongation
XX and modification. The minimal PKS module is typified by module 3 which
XX contains a ketosynthase domain, an acyltransferase domain, and an acyl
XX carrier protein domain. These three enzyme activities are sufficient to
XX activate the 2-carbon extender unit and attach it to the growing
XX polyketide molecule. Additional domain which may be present include a
XX ketoreductase domain, an enoylreductase domain, and a dehydratase domain.
XX The PKS polypeptide is useful for preparing epothilone in large
XX quantities. Epothilone is a polyketide antibiotic that also has
XX antitumor activity.
XX
XX Sequence 751 BP; 103 A; 220 C; 302 G; 126 T; 0 other;
SQ
Query Match 54.1%; Score 54.6; DB 21; Length 751;
Best Local Similarity 71.3%; Pred. No. 4.8e-06;
Matches 72; Conservative 0; Mismatches 29; Indels 0; Gaps 0;
OY 1 AGCGGCTGGAGGCTGGCTCCGAGCGAGGCTGTCAGACTCTCTCGCCGACGCCAC 60
DB 130 AGCTGCTGCCCGGGGCTGGCGAGACGCTGCTGCAAGACCTCTTCTCCAGGCTGAC 189
OY 61 GCGCTGGGGTGGAGCGAAGGCTGGCCATGCTCTCTCAAA 101
DB 190 GCGCTCATCTGTGTCGGAAGGCTGGCGGATGCTGTCTGAA 230
RESULT 11
AAT68715
ID AAT68715 standard; DNA; 15872 BP.
XX
XX AAT68715;
XX
DT 01-SEP-1997 (first entry)
XX
```

```
DE Streptomyces venezuelae polyketide synthase vep ORF1.
XX
XX Polyketide synthase; polyhydroxyalkanoate monomer synthase;
KW polyhydroxybutyrate; biodegradable polymer; vep gene;
KW metabolic engineering; ss.
XX
XX Streptomyces venezuelae.
XX
FH Key Location/Qualifiers
FT CDS 20..13912
FT /tag= a
FT CDS 14056..14136
FT /tag= b
FT CDS 14148..15827
FT /tag= c
XX
XX WO9722711-A1.
XX
XX 26-JUN-1997.
XX
XX 18-DEC-1996; 96WO-US20119.
XX
XX 19-DEC-1995; 95US-0008847.
XX
XX (MINU ) UNIV MINNESOTA.
XX
XX Sherman DH, Williams MD, Xue Y;
XX
XX WPI; 1997-341701/31.
XX
XX P-PSDB; AAM19629-30 AND AAM00918.
XX
XX Expression cassettes for production of polyhydroxyalkanoate(s) -
XX provide wide range of biodegradable polymers for medical or
XX industrial use
XX
XX Claim 54; Fig 23; 91pp; English.
XX
XX Streptomyces venezuelae vep ORF1 (AAT68715) comprises the polyketide
XX synthase (PKS) gene cluster encoding a polyene of 12 carbons (see
XX also AAM19629-30 and AAM00918). It contains 5 PKS modules, with a 5'
XX loading module and a 3' end domain. Each of the sequenced modules
XX includes a keto-ACP, an acyltransferase, a dehydratase, a keto-
XX reductase and an acyl carrier protein domain. The gene cluster was
XX cloned using a heterologous hybridisation strategy from a genomic
XX DNA library. A novel expression cassette encoding the first module
XX from the vep gene cluster and module 7 from the Streptomyces ty1P
XX gene cluster has polyhydroxyalkanoate (PHA) monomer synthase
XX activity and can be used for PHA prodn. in host (esp. insect) cells
XX for use as a biodegradable polymer.
XX
XX Sequence 15872 BP; 2093 A; 6313 C; 5503 G; 1963 T; 0 other;
SQ
Query Match 53.9%; Score 54.4; DB 18; Length 15872;
Best Local Similarity 72.9%; Pred. No. 6.2e-06;
Matches 70; Conservative 0; Mismatches 26; Indels 0; Gaps 0;
OY 3 CCGGCTGGAGGCTTGGCTCCGAGGAGCGTGGCAAGACTTCTCGCCGACGCCAGG 62
DB 9136 CCGGAGGAAGGCGGCGCGCGCGCTGTCGCAAGGCGTTCTGCGCGCGCGAGG 9195
OY 63 CGTGGGTGGAGCGAAGGCTGGCCATGCTCTCTGCT 98
DB 9196 CACCGGCTGGCGCGGAGGCGTGGGCTGCTCTCTCT 9231
RESULT 12
AAZ87283
ID AAZ87283 standard; DNA; 15872 BP.
XX
XX AAZ87283;
XX
XX 05-JUN-2000 (first entry)
XX
```

DE	XX	S. venezuelae vep ORF 1, SEQ ID NO:1.	
KM	XX	Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;	
KM	XX	nemiahymycin; nabhomycin; polyhydroxylkanonate monomer synthase;	
KM	XX	biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;	
KM	XX	chronic obstructive pulmonary disease; respiratory inflammation;	
KM	XX	hypercholesterolemia; crop protection agent; ds.	
OS	XX	Streptomyces venezuelae ATCC15439.	
FH	XX	Key	Location/Qualifiers
FT	XX	CDS	20..13912
FT	XX		/*tag= a
FT	XX	CDS	/*product= "vep ORF 1 amino acid sequence #1 (AAV77177)"
FT	XX		14056..14151
FT	XX		/*tag= b
FT	XX	CDS	/*product= "vep ORF 1 amino acid sequence #3 (AAV77199)"
FT	XX		14167..15827
FT	XX		/*tag= c
FT	XX		/*product= "vep ORF 1 amino acid sequence #2 (AAV77178)"
PN	XX	WO200000620-A2.	
PD	XX	06-JAN-2000.	
XX	XX		
PF	XX	25-JUN-1999;	99WO-US14398.
PR	XX	26-JUN-1998;	98US-0105537.
PA	XX	(MINU) UNIV MINNESOTA.	
PI	XX	Sherman DH, Liu H, Xue Y, Zhao L;	
XX	XX	WPI; 2000-160679/14.	
DR	XX	P-PSDB; AAV77177, AAV77178, AAV77199.	
XX	XX	Desosamine and macrolide biosynthetic gene clusters, useful for, e.g.	
PT	XX	synthesis of methymycin and pikromycin -	
PS	XX	Example 3; Figure 23; 438pp; English.	
XX	XX	The invention relates to an isolated and purified nucleic acid segment	
CC	XX	comprising a desosamine biosynthetic gene cluster, a fragment or its	
CC	XX	biologically active variant, where the nucleic acid sequence is not	
CC	XX	derived from the eryc gene cluster of Saccharopolyspora erythraea or	
CC	XX	Streptomyces antibiotics. The invention also relates to a macrolide	
CC	XX	biosynthetic gene cluster, or fragments thereof. The macrolide	
CC	XX	biosynthetic gene cluster encodes proteins which synthesise methymycin,	
CC	XX	pikromycin, nememiahymycin, nabhomycin or a combination of these	
CC	XX	compounds. Recombinant or augmented cells comprising the desosamine	
CC	XX	and/or macrolide biosynthetic gene clusters are useful for the	
CC	XX	production of biologically active macrolides. The macrolide biosynthetic	
CC	XX	proteins are useful for synthesis of methymycin, pikromycin,	
CC	XX	nememiahymcin and nabhomycin. The alternative termination of polyketide	
CC	XX	synthesis may be useful to prepare novel antibiotics and	
CC	XX	polyhydroxylkanonate (PHA) monomers. The compounds produced by the	
CC	XX	recombinant host cells are useful as biopolymers, e.g., in packaging or	
CC	XX	biomedical applications, to engineer PHA monomer syntheses or to prepare	
CC	XX	immunosuppressants, agents, such as chemotherapeutics,	
CC	XX	immunosuppressants, agents to treat asthma, chronic obstructive pulmonary	
CC	XX	disease as well as other diseases involving respiratory inflammation,	
CC	XX	cholesterol-lowering agents or macrolide-based antibiotics which are	
CC	XX	active against a variety of organisms, e.g., bacteria, including	
CC	XX	multi-drug resistant pneumococci and other respiratory pathogens, as well	
CC	XX	as viral parasitic pathogens, or as crop protection agents (e.g.,	
CC	XX	fungicides or insecticides) via expression of polyketides in plants. The	
CC	XX	present sequence represents a Streptomyces venezuelae ATCC 15439 DNA	
CC	XX	sequence, designated vep ORF 1 in the specification, which actually	
CC	XX	contains 3 open reading frames, which encode proteins AAV77177-v77178 and	
CC	XX	AAV77199. The vep ORF 1 protein is defined in the specification as a PHA	
CC	XX	monomer synthase.	
XX	XX	Sequence 15872 BP; 2088 A; 6304 C; 5513 G; 1967 T; 0 other;	

Query Match	Similarity	53.9%	Score	54.4	DB	21	Length	15872	
Best Local	Similarity	72.9%	Pred.	No. 6.2e-06					
Matches	70	Conservative	0	Mismatches	26	Indels	0	Gaps	0
Qy	3	CCGCGTCGCGAGCCCTGGCTCCCGACGAGCGTCAAGACTTCTCGCCGCGACGCGACG	62						
Db	9136	CCGGCAGAAAGGAGGCGTGGCCGCGACGCGCGTGCAGAGCGTTCGCGCGCGCGACG	9195						
Qy	63	CGTGGGCTGAGACGAAGGCTGCGCATCTCTGCT	98						
Db	9196	CACCGGCTGGCGGAGGCGTGGGCTGCTGCTCT	9231						
RESULT 13									
AAAT80413	ID	AAAT80413	standard	DNA	43280	BP			
XX	AC	AAAT80413							
XX	DT	27-FEB-1998	(first entry)						
XX	DE	Tylactone synthase gene cluster.							
XX	KW	Tylactone synthase gene cluster; tylG gene; multifunctional protein;							
XX	KW	polyketide; tylactone synthesis; antibiotic; tylosin; ss.							
OS	Streptomyces fradiae.								
XX	FX	Location/Qualifiers							
XX	FT	816..1423							
XX	FT	/tag= a							
XX	FT	/transl_except= (pos: 816..818, aa: Met)							
XX	FT	/note= "ORP1 encodes protein shown in AAW22601"							
XX	FT	14351..19945							
XX	FT	/tag= b							
XX	FT	/transl_except= (pos: 14351..14353, aa: Met)							
XX	FT	/note= "ORP2 encodes protein shown in AAW22602"							
XX	FT	20010..31199							
XX	FT	/tag= c							
XX	FT	/transl_except= (pos: 20010..20012, aa: Met)							
XX	FT	/note= "ORP3 encodes protein shown in AAW22603"							
XX	FT	31232..36067							
XX	FT	/tag= d							
XX	FT	/note= "ORP4 encodes protein shown in AAW22604"							
XX	FT	36249..41774							
XX	FT	/tag= e							
XX	FT	/note= "ORP5 encodes protein shown in AAW22605"							
XX	EP	EP791655-A2.							
XX	PD	27-AUG-1997.							
XX	PF	19-FEB-1997; 97EP-0301056.							
XX	PR	22-FEB-1996; 96US-0012078.							
XX	PA	(ELIL) LILLY & CO ELI.							
XX	PI	Dehoff BS, Kuhstoss SA, Rosteck PR, Sutton KL;							
XX	DR	WPI; 1997-418046/39.							
XX	DR	P-PSDB; AAW22601-W22605.							
XX	PT	DNA encoding Streptomyces fradiae tylactone synthase domain - for							
XX	PT	production of tylosin-related polyketide compounds							
XX	PS	Claim 2; Pages 8-66; 220p; English.							
XX	XX	This sequence represents the tylactone synthase gene cluster of the							
CC	CC	invention. This sequence is also referred to as the tylG gene, and was							
CC	CC	isolated from Streptomyces fradiae. This sequence encodes multifunctional							
CC	CC	proteins which direct the synthesis of the polyketide tylactone, isolated							

CC from Streptomyces fradiae. Tylactone is the basic building block of the
CC antibiotic tylosin. The DNA sequence can be modified so as to alter the
CC type of carboxylic acids incorporated, the number of carboxylic acids
CC incorporated and/or the post-condensation reactions performed, thereby
CC resulting in novel tylosin-related polyketides.

XX Sequence 43280 BP; 5269 A; 15187 C; 16651 G; 6173 T; 0 other;

Query Match 53.9%; Score 54.4; DB 18; Length 43280;

Best Local Similarity 72.9%; Pred. No. 6.4e-06; Matches 70; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 3 CCGGCTGCGAGGCTGTGCTCCGACGAGGTGCAAGAGCTTCTGCGCGACGCGACGG 62

Db 36998 CCGGCAAGAAGGCGCTGGCGCGGACGCGCGTCTGCGCGCGCGCGACGG 37057

QY 63 CGTGGGTGAGCGAAGCGTGGCCATGCTCTGCT 98

Db 37058 CACCGCTGGCGCGGCGGTGCGCTGCTGCT 37093

RESULT 14

AAD17184 AAD17184 standard; DNA; 65140 BP.

XX AAD17184;

XX 29-NOV-2001 (first entry)

DE Streptomyces noursei nys1 DNA of nystatin PKS gene cluster.

KW Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;

KW antifungal; antibiotic; nys1; ds.

OS Streptomyces noursei.

XX Location/Qualifiers
FH complement (1..1035)

FT /tag= a
FT /product= "NysD2 partial protein"

FT /note= "CDS does not include stop codon"

FT /tag= b
FT /product= "NysD1 protein"

FT /tag= c
FT /product= "NysA protein"

FT /tag= d
FT /product= "NysB protein"

FT /tag= e
FT /product= "NysC protein"

FT /tag= f
FT /product= "NysE protein"

FT /tag= g
FT /product= "NysR1 protein"

FT /tag= h
FT /product= "NysR2 protein"

FT /note= "CDS does not include start codon"

FT /tag= i
FT /product= "NysR3 protein"

FT /tag= j
FT /product= "NysR4 (short) protein"

FT /note= "CDS does not include start codon"

FT /tag= k
FT /product= "NysR5 protein"

FT CDS

FT CDS

FT CDS

FT CDS

FT /note= "CDS does not include start codon"

FT complement (62551..63615)

FT /tag= l
FT /product= "ORF2 protein"

FT /note= "CDS does not include start codon"

FT /tag= m
FT /product= "ORF1 protein"

XX WO200159126-A2.

XX 16-AUG-2001.

XX 08-FEB-2001; 2001WO-GB00509.

XX 08-FEB-2000; 2000GB-0002840.

XX 10-APR-2000; 2000GB-0008786.

XX 14-APR-2000; 2000GB-0009387.

XX (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.

XX (SMTF) SINTER STIPELSEN IND TEK FORSK.

XX (ALPH-) ALPHARMA AS.

XX (SINV-) SINVENT AS.

XX (DZIE/) DZIEGLEWSKA H.

XX (ZOTC/) ZOTCHEV S B.

XX (SEKU/) SEKUROVA O N.

XX (FJAE/) FJAEVYIK E.

XX (BRNU/) BRAUTASET T.

XX (STRO/) STROM A R.

XX Zorchev SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AR;

XX Valla S, Ellingsen TE, Sletta H, Gulliksen O;

XX WPI: 2001-557614/62.

XX P-PsDB: AAE10125, AAE10126, AAE10127, AAE10128, AAE10129, AAE10130,

XX AAE10131, AAE10132, AAE10133, AAE10134, AAE10135, AAE10136, AAE10137.

XX New nystatin polyketide synthase polynucleotides and polypeptides,

XX useful as antibiotics and antifungals -

XX Claim 2; Page 116-151; 266pp; English.

XX The present invention relates to the cloning and sequencing of the gene

XX cluster encoding a modular type I polyketide synthase (PKS) enzyme

XX involved in the biosynthesis of the macrolide antibiotic nystatin.

XX The nystatin PKS is useful as antifungal antibiotics. The present

XX sequence is a Streptomyces noursei nys1 DNA of nystatin PKS gene cluster.

XX Sequence 65140 BP; 8270 A; 25171 C; 22273 G; 9426 T; 0 other;

XX Query Match 53.9%; Score 54.4; DB 22; Length 65140;

XX Best Local Similarity 72.9%; Pred. No. 6.5e-06;

XX Matches 70; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 3 CCGGCTGCGAGGCTGTGCTCCGACGAGGTGCAAGAGCTTCTGCGCGACGCGACGG 62

Db 34048 CCGGCAAGAAGGCGCTGGCGCGGACGCGCGTCTGCGCGCGCGACGG 34107

QY 63 CGTGGGTGAGCGAAGCGTGGCCATGCTCTGCT 98

Db 34108 CGTGGGTGAGCGAAGCGTGGCCATGCTCTGCT 34143

RESULT 15

AAD17186 AAD17186 standard; DNA; 125401 BP.

XX AAD17186;

XX 29-NOV-2001 (first entry)

DE Streptomyces noursei nystatin PKS gene cluster DNA.

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 6, 2002, 11:51:49 ; Search time 443.625 Seconds

(Without alignments)
6625.825 Million cell updates/sec

Title: US-09-724-876-2_COPY_32070_32170

Perfect score: 101

Sequence: 1 agcgcgtcgcagcctgcgc.....tgcgcattctctctctcaaa 101

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapect 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:*

1: gb_ba:*

2: gb_htg:*

3: gb_in:*

4: gb_cm:*

5: gb_ov:*

6: gb_pat:*

7: gb_ph:*

8: gb_pl:*

9: gb_pr:*

10: gb_ro:*

11: gb_sts:*

12: gb_sy:*

13: gb_un:*

14: gb_vl:*

15: em_ba:*

16: em_fun:*

17: em_hum:*

18: em_in:*

19: em_mu:*

20: em_om:*

21: em_or:*

22: em_ov:*

23: em_pat:*

24: em_ph:*

25: em_pl:*

26: em_ro:*

27: em_sts:*

28: em_un:*

29: em_vl:*

30: em_htg_hum:*

31: em_htg_inv:*

32: em_htg_other:*

33: em_htg_mus:*

34: em_htg_pln:*

35: em_htg_rtd:*

36: em_htg_man:*

37: em_htg_vrt:*

38: em_sy:*

39: em_htgo_hum:*

40: em_htgo_mus:*

41: em_htgo_other:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	101	100.0	58733	1 AF217189	AF217189 Sorangium
2	101	100.0	68750	1 AF210843	AF210843 Sorangium
3	101	100.0	68750	6 AR193029	AR193029 Sequence
4	101	100.0	68750	6 AR199551	AR199551 Sequence
5	101	100.0	68750	6 AR199559	AR199559 Sequence
6	101	100.0	68750	6 AR199567	AR199567 Sequence
7	101	100.0	68750	6 AR201097	AR201097 Sequence
8	101	100.0	68750	6 AR208671	AR208671 Sequence
9	101	100.0	71989	6 AR172664	AR172664 Sequence
10	59.8	59.2	104326	1 AF411573S4	AF411573S4
11	59.2	58.6	3185	1 AF016585	AF016585 Streptomy
12	58	57.4	41097	1 SNA13222	SNA13222 Streptomy
13	57.8	57.2	20394	6 AX067996	AX067996 Sequence
14	57.8	57.2	20394	6 AX067996	AX067996 Sequence
15	57.8	57.2	27522	1 AB070942	AB070942 Streptomy
16	57.8	57.2	84985	1 SNA278573	AJ278573 Streptomy
17	57.6	57.0	672	6 AX153706	AX153706 Sequence
18	57.6	57.0	104326	1 AB070940	AB070940 Streptomy
19	56.6	56.0	78210	1 AB070949	AB070949 Streptomy
20	56.4	55.8	24225	1 SC2C4	AL512902 Streptomy
21	56.4	55.8	30000	6 AX250261	AX250261 Sequence
22	56.2	55.6	32870	1 AF007101	AF007101 Streptomy
23	56.2	55.6	33529	6 AR166425	AR166425 Sequence
24	56	55.4	113193	1 AF357202	AF357202 Streptomy
25	55.8	55.2	14402	1 AF098795	AF098795 Pseudomon
26	55	54.3	30000	6 AX250262	AX250262 Sequence
27	54.8	54.3	10910	1 AX024383	AX024383 Sequence
28	54.8	54.3	10910	6 AX024276	AX024276 Sequence
29	54.8	54.3	39314	1 SGR300302	AJ300302 Streptomy
30	54.8	54.3	47981	1 AF263245	AF263245 Micromono
31	54.8	54.3	47981	6 AX112026	AX112026 Sequence
32	54.6	54.1	22838	1 AF081920	AF081920 Pseudomon
33	54.4	53.9	30000	6 AX250263	AX250263 Sequence
34	54.4	53.9	43280	1 SFU78289	U78289 Streptomyce
35	54.4	53.9	65140	6 AX211705	AX211705 Sequence
36	54.4	53.9	82746	1 AF453501	AF453501 Actinosyn
37	54.4	53.9	123580	1 AF263912	AF263912 Streptomy
38	54.4	53.9	125401	6 AX211739	AX211739 Sequence
39	54.2	53.7	67523	1 SCU24241	U24241 Sorangium c
40	53.6	53.1	66808	1 SAU421825	AJ421825 Stigmatel
41	53.2	52.7	16767	6 AX089464	AX089464 Sequence
42	53.2	52.7	26195	1 SC1G7	AL591083 Streptomy
43	53.2	52.7	30000	6 AX250263	AX250263 Sequence
44	53.2	52.7	50000	6 AX089417	AX089417 Sequence
45	53.2	52.7	50000	6 AX089420	AX089420 Sequence

ALIGNMENTS

RESULT 1

AF217189

LOCUS

DEFINITION

AF217189 Sorangium cellulosum putative transposase gene, partial cds;

putative transposase gene, complete cds; epothilone biosynthesis gene cluster, complete sequence; putative membrane protein gene, complete cds.

ACCESSION

AF217189

VERSION

AF217189.1 GI:7453554

KEYWORDS

SOURCE

Polyangium cellulosum.

ORGANISM

Polyangium cellulosum

Bacteria; Proteobacteria; delta subdivision; Myxobacteria; Myxococcales; Sorangineae; Polyangiaceae; Polyangium.

REFERENCE

1 (bases 1 to 58733)

Pred. No. is the number of results predicted by chance to have a

AUTHORS	Tang,L., Shah,S., Chung,L., Carney,J., Katz,L., Khosla,C. and Julien,B.
TITLE	Cloning and heterologous expression of the epoethlone gene cluster
JOURNAL	Science 287 (5453), 640-642 (2000)
MEDLINE	2015553
PUBMED	10649995
REFERENCE	2 (bases 1 to 58733)
AUTHORS	Julien,B., Shah,S., Ziermann,R., Goldman,R., Katz,L. and Khosla,C.
TITLE	Isolation and characterization of the epoethlone biosynthetic gene cluster from Sorangium cellulosum
JOURNAL	Gene 249 (1-2), 153-160 (2000)
MEDLINE	20293058
PUBMED	10831849
REFERENCE	3 (bases 1 to 58733)
AUTHORS	Julien,B
TITLE	Direct Submission
JOURNAL	Submitted (16-DEC-1999) Kosan Biosciences, Inc., 3832 Bay Center Place, Hayward, CA 94545, USA
FEATURES	Location/Qualifiers
source	1..58733

```

CDS
>1..992
/note="orfA"
/godon_start=3
/transl_table=1
/product="putative transposase"
/protein_id="AAF62878.1"
/db_xref="GI:745355"
/translation="VRGHEAFADGCGVRLVLDYDNKNAVERHGDIAIRPHTPLIAL
SADYREPRPAVAVRGNCKGKVERAIRVRRGFGFEARVADGLNRRATGTSAAAL
DSRWEDARVYRQAGFDERSVLLRHPTPPDHERVEVEGKPTPIYAFDNDIVSPP
DRIKRLVADLSQVRIADNQIVATHVNRSDQQQLDEQHIQLRDVDEKRRREHSH
GLDRIARARSQAQLRIVAERGDGVGSAIRLLQLDLVGAALAEELVLEVRDIIL
HIGVAVQVIDIRRSERHRLPPVSIPTVGEHAALVWTPHSLTYDALKKDPTP"
989..1501
CDS

```

```

/product="putative transposase"
/protein_id="AAF62879.1"
/db_xref="GI:7453556"
/translation="MTDITPEYRKDRLSLGLFGLACMEQLADKPWLRVLALEERF
RHKSLERLRLNSRAARKPMTDFDSSMPKIIDDEAVDLYDSRAOILFEVVRKDY
AQKPLLSTNKAFAADMGVFPFACVYTLVDRLYHRAVEITEASTYRLKAELNAT
RTKDKRIKTH
1998..6263
/gene="epoa"
/gene="epoa"
1998..6263
/gene="epoa"
/feature="type I polyketide synthase; epothilone PKS loading
module"

```

```

/protein_id="AAFG5880.1"
/db_xref="gi:7453557"
/translation="MADPRIRAEADPIATVAGACRLPGYIDISGEWTLLEGRDVI
GQVPERMDAAWFDPLDADGKTPVTRASLSQVACDAEAFGISPEALRMDPGTH
LILECVAERMDENAAIASPSALVGTGVEVFGISGPEVEMASIDHGGHGLPMH
PSVAGASRTYVLGRLGPCVAVDTAVSSSLVHLACLOSREGCESTALAGVSSILG
STLWLSRTALATDGRCKAFSAEADGREGGCAVYLKTLSCARADGDLIIVLGG
SAIHDDCASSSLTYVNSSQSEIYLAKRALADGCAASSGYEANGTGTTLDDPLEIQM
LNAVYGRDVAIPLTGVSVTNLGHPETASITGLKLVNLSLHGQIPALHAAQOLAL
PRISWGLDIRVATPRAPVPMQMTNPRAGVSSFGMSGNAVHVEEAAPCEAPAPR
PRAELVLSATATAALDAHLAHRDLHETYSQCLGDVAFELATTRSMERHLAVAPR
SSEELRALDAAAGOTPPGVGRVIAIDSSRCKLAEFLGCGAOTGLGMRGLDVPWAPR
REARDICLHENOELDPLREVMMAAPRAEYUALLDOTAFOPALTFEYVALATMS
MGCEPILVAGHSISGELVACVAGVSEDAVETUAAKRRLQALPAGGAVSTAPERS
DYAAVAPHASVSIANNGPDQVILKAGQPVHAIILAAKARAKTALVSHAFHSIPR
PLAMPLEAFGEVAAVESYRRPSIVLVSPLKCGKTEGVSSGQVWRARVAPRADOQ
VKALHAAGAGTVEVGRSTLTGLVPLCLPDRKALLAASSRAGDEBAPVTEALGGLM

```

gene
CDS

CDS

gene
CDS

CDS

```

AVGVLSGMSAASGAGRRVLPVPMQREKVRIDIRKADDAARCGRRAGACHDEVC
GAVGVGMSAASRLDPHGGRRREKVEYEAAGDRPFLDIEDEGVDRILRATERBEG
LGEVEIANDAGSIFNDVOLALCMEVDVDPGKPNPPLLGGGCGRIVAAEGEVNGI
VGEVYIADSGAGVATVQWASQVAGVEVATGATGEKRAYLESLGVAVSOSRDRFV
SERVYLIIAAGAGVLAAGVAAQVAGVEVATGATGEKRAYLESLGVAVSOSRDRFV
NDSVAMGEGEVDVNLNSLGEELIDFNLISAGVPELGRDVCYANDQAFPSMO
LAFSTIVDIDGMMLEPARRARLEFEELGLIIAGVPTPIATLIVANDAFPSMO
AOHGLKVLITGDEPVOIRPTHAGSGPSTGRDILDLRLASAPAAALAEFLRQ
VSQVLPREIKVGEALFTGLGDSILMAELNRIEASIKLSIKLSTPLSTSPNTALT
ONLIDALATLALSERVAENLRAGVDSDFVSGAQDQDEWIIAL"
/gene="epoc"
6260. 10492
/gene="epoc"
6260. 10492
/gene="epoc"
6260. 10492
/module 1"
/codon_start=1
/transl_table=11
/product="Epoc"
/protein_id="AA662861.1"
/db_xref="GI:7453558"
/transl_table="metazoa"
KSTIITLIRKLPALPESVLPABRVPPFLDTGOSVLMGRGATVYSGTHARFED
CTIDIVARLSAPKPVYARHMLRATLTDMMQVIEPYVDADIELDRIGDRSTERRA
RLVSLRQMSHRITVDPREPIYVAVVRLDEOQTULSLDIANDGSLSTIKPKDL
SPYDEPSELVELELSDRVYLLSERRKSEHQSMDIKWRVAELPPMLPKAD
PSTLEIREFRTEEMDPSDSMSLKQVBERGLPVGVIILAAFSVIGRMSASRFL
NITTELRVPHVRNLTGTGFTSMVLDDITVYRDSFEQRAKRIOEOLMMSHDCVS
GIEVQREARVLRIGORALEPVLVTSALNOQVAYVSLQSTGTPYVTSOTPLRDL
LHESNDQVILVANDIYGVPEPRLDMLAEVAFRLTEPMSRSGSLPAQOLE
ARASANEPLNSLPHLILGLFARVQRLPMOLAVASAKRTIEPDRSRRGRARL
EOGARPTVLAAYMEKGEVYAVLAVLESGAAYPIDADPAERIITLHDGCVKIV
LQGMIDHKAISWPEIGLRLISGAGVSGGIDLPMPITQTSDDLAVVITYGSGTLEK
VGMIDHGAQVAVTILIDNERELTSPGVRVGLSLSFEDLSVDYVGLIIAAGTIVPRA
SKLDPHMAHMLIEREVYVYVNSVPMVLEHSEGPDSIARSLRSLSGMIVP
GLPELDAIRGVSVISLGGATGASIMVIGYVAVNDSPMSIPIGRLNQTGFVLD
EALRPVPMVPGQVITGGVLAGYRDRDEKTRNSFVHPETGELVYTGDLGVLPD
GNTERMEEDQIKLRTVRELDELETLSHPNRDVAIVPGANDAKMLLAVP
BGTERRAEADQSLKTERVDARHAHAKADGLSDGRVQFKLARHGLRDLDKGVND
TGLVPRREGADQSLKTERVAKRSVTFLEAIPVEVGRFLSCISVPEPGALPEFRPSAG
STVQVYAAKSGARLSEGVDEGYUHPENHLLVSHQIERGHVQNDVDEEAA
GLTFQVDRIDAIESVGLSRECLILEAGYMLLMBOAQNIGCVQGFDEYEA
PVLDRHSDVYVHMLGGRVDPQVQVPTGLDSSFRPATTRTGRAPRGNDQFALD
FLRLKEPIVMPVFEYELDLPLTSGKVDKRALERKQTSPRSHGTARADLEI
LAVARVLESLDLEVGQDSVFDGATSIHIVMRSLLOKRLDREIATITELFOYPNLSS
LASGLRRTSKLEQVPRMMDRVARRRKRRRS"
10639. 16137
/gene="epoc"
10639. 16137
/gene="epoc"
10639. 16137
/notice="type I polyketide synthase; epothilone PKS module
2"
/codon_start=1
/transl_table=11
/product="Epoc"
/protein_id="AA662862.1"
/db_xref="GI:7453559"
/transl_table="metazoa"
ALNAGVDPPLAVIDPNVYRAGVSLVDFEDFAFGISPRAEALMDPOHRTFMCACAE
ALNAGVDPPLAVYRAGVSLVAGAMMSYLTLSNHEPRAAPRGAFQTLIGDKQYLAT
HAYVRLMLRGRSISVDTAGTSILVAYHLCMSLDRDEDMALAGITVIRIDHRAGVY
AAEGISIPEDHSGCFADKAKAGTIMGSCGVLYKPLDRLSDGQVRVYLIIASLTND
GARIGITPASHGEVQALAMEALVALIGVARSIVQIEHNGTILGDAIEGASVPRV
DRDSVSTRSCLISVKGIGHESAGSINGILKTVLALHNOQLPFLNFPESNPISDF
ASFPYVTSIKMWNQSPRRRAGVSEFICGTNAHYVLEAPRAAKPRASASAL
FVNSAKSAALDAAARLRHDLHOGVLSGVNATSLTSPBMRILAAAPSRAL
REGIDAAARQTPGVAVRGCSGVNVPVPOQSGQWGMKOLALMEVGFHAL
SACRADAOGAGSILAEIADLADGSSQIERIDVQVFLFALVALIAMSQVGPV
VIGSHMEVAVNSHVAAGLSLEDAVILICRSRLIRISGQEMAVTELISLAEADAR
GIDRVSVAIVANSRPSLGEPAVIGELSSILAKGVCSRRVAVDVAHSPQVDRP
EDLLALAGLRPPRAAVPBRSTYTGAHVAPELGANTYMMNDRQVRAEYVDAQLOG
PHGILFEMSPSHRPLITVTSVEEMRAARAGAAVGSIRQODERPMALDELGAHQGP
GPGVLEPPAGGRPLPITVPMQREKVRIDIRKADDAARCGRRAGACHDEVCOTLS

```



```

CDS
    /product="unknown"
    /protein_id="AA26907.1"
    /db_xref="GI:6724241"
    /translation="MDKPIGRCAIAEGYIPGSGNGPEPOMTSHETACILNASDRDA
    OVAITVFSDDPAGPRVTPVAPRRTRHVRFDLLEPEIRIPDPTQYASVIESDAIV
    OHTRLDSQAEALLSTIAYTDRE"
    6236..6675
    /note="ORF5"
    /codon_start=1
    /transl_table=11
    /product="unknown"
    /protein_id="AA26908.1"
    /db_xref="GI:6724242"
    /translation="MKHVDGRRFRGRIHTGLLASMALAGCGPSEKTVQGTSLAP
    GADAVATADVPDPAATRLAVDVHLSPPERLEAGSERPVVWQRPSPSPMRVGVLD
    YNADSRCKLAETTVYANFELLIAEKQSSQSPSSAAVIGTFSV"
    7610..11875
    /gene="epoa"
    /note="epoa"
    /gene="epoa"
    /note="EPOS A"
    /codon_start=1
    /transl_table=11
    /product="polyketide synthase"
    /protein_id="AA26919.1"
    /db_xref="GI:6724253"
    /translation="MADRIERAEDPIAIVGASCLPGVIDLSGFWTLLEGSDTV
    GVPAERMDAAMFDDPDPAKGTPTVTRASFSDVACDAPSGISPRALMDPAHR
    LILECYMALEMAAIAAPSAIVGTGFGISPEYEAALPQATASAEIDANGCTM
    PSVCGRISTYALGLGPCVAVDTATSSSLVANHLCOSIRSECGTALAGYSIMLSP
    STLWLSKTRALARDCRCARFADGFGGCGCAVVILKLSGARABGDRLLAIVRG
    SAINHDGASGGLTPNGSSOEIVLRALADAGCAASVGVANGHTGLGPIDIAQ
    LNAVYGLRDVATPLIGSVKTLNGHPEASGITGLKLVLSLQHQIIPAHIAQALN
    PRISGDRRLTYTRATPMPDMNTPRAGVSGFGSGTNAHVLEAPAACTPPAPE
    RPAELVLSARTASALDAQAARLDHLETPSQCQDVAFSLATRSAMENHLAAT
    SREGRLALDAAGCQTSFGAVRSTADSSRCKLALFTGQGNQITGMGEGYLDVMSAF
    REAFDLVRLFNOELDRPLREYMAEPASVDALLDQTAFTQPALFTEGYLDVMSAF
    MGVEPELVAGHSIGELVACVAGVPSLEDVAVLVARGLMQLPAGAMVSIEMPEA
    DVAAVAVPAAASVSIAAVNAPOVVITAGGQVHAIAAAMARGARTKALHVSFHS
    PLMAPLFEFGFVAVSVGRPSIVLSVLSGKACTDENVSPGVVHRAHREVRADG
    VKALHAGACGTFVSEVSGKSTLLGLVPACMPDRAPALLASSRAGREPRATVLEALGIM
    AVGLVSNAGLFPSCGRKVPPLTTPMQREKIVITDKADDAAGDRRAGAGDEVEEG
    GAVRGDRRSARLDHPPESGRERREVEAAGDFPLEIDEPEVLDHLVITERAPAG
    LGVEIVADVAGLSTFNDYDOLALGMVDDLPKGNPRLLEIGECACRIAYAVGVNGLV
    VGQPVIALSAGAFATHTTSAALVLRQALSAIPAAMPVAYLGVARYLDRIDIAIOP
    GERVLIAHATGCVGLAAVOMAGVAGVHATGTPREKRAYLESIGVRYVSDSRAPV
    ADVRAWTGCEGVYDYNLSGELIDKSNLRSRFPYELGRDCYADNOGLDFELR
    NLSESVDLRGMMLEPRARVRLLELELGLIAGVFTPPPIATLPIARVADFRMAQ
    AQHLGLVLTGDPVQIRITPHAGAGSTGDRDLDRLASAAPARAALAEAFRLQ
    VSOVLRTEIKVGALEFTRLGMDSLMVELNRIEASIKLSTTFLSTSPNIALLA
    ONILDALATLSLEVAENLRAGVQNDPVSAGADQWEIIAL"
    7640..8920
    /gene="epoa"
    /note="Region: beta-ketoacyl synthase/malonyl-CoA
    decarboxylase"
    9236..10201
    /gene="epoa"
    /note="Region: acetate acyltransferase"
    10529..11428
    /gene="epoa"
    /note="Region: enoyl reductase"
    11549..11764
    /gene="epoa"
    /note="Region: acyl carrier protein"
    11872..16104
    /gene="epop"
    11872..16104
    /gene="epop"
    /note="EPOS P: contains condensation/heterocyclization,
    cysteine adenylation, putative NAD(P)H oxidase, and
    peptidyl carrier domains"
    /codon_start=1
    /transl_table=11

```

```

    /product="nonribosomal peptide synthetase"
    /protein_id="AA26925.1"
    /db_xref="GI:6724259"
    /translation="MTINOLNEHEHOGIKTLAODERLOIOAPKNALPNLLARISEH
    KSTILTMLORLPAESIYPAPAEHRAPEPLDIOESVLTGRTGAFTVPSGTHAVREVD
    CTDLDPRLSRARFRKVVARHDMLRHNTPLDMQYTERVVDIDIELDRGRDSTREA
    RLVSRLDMSHRTIDYTERPPLYHVAARLDERQYRLVSLIDINDLSSLSITRDLW
    SEYEDPETSLPYLEISYRDYVLALESRSKSAHOSMDYMRRLAELEPPPLPMKAD
    PSTLKEIRFRHTEQMLPSDSWGLRKLRGEGGLPTGYILAFSASVIGRMSASPRFL
    NITTLERLPHVRVNDITGDFTSMVLIDITTRDKSPFORAKRIOELAMDDVDVS
    GTEVOREARVILGORGALFPVVLTSALNOGVYVTSIRGTEPVTYSTOTPOLLOH
    OLYEHNDGLVIANMDYVDGFPFDDLDLDEAVYVTRLTLEPMEGEOYRCSLPQOLE
    ARASANATNALISEHTLGLTPRARVEQLPQDLAVVSAKTLTTELSRRSRRLARLR
    EOGAPNTLVAVMEKGEOVAVVAVLAVESAAVYPIDADI.PARITHLIDHGVKLV
    LTQPLDLKLSMPGLOHLVSEAGVEGDGQPPMPPIOTSDLAIVYVTSGSGLPK
    GVMIDHRCGAVNTIIDINRFETIGPDRLVLSSEFDSVYDGLIAGGTIVVPA
    SKLRDPAHMAVELIEREKTYVNSPALRMVLEHREGRDVSILASIRLSLSGMIWY
    GLPGLQAIRVSVYISLGAATENASTIGTPVRVNDLSMASITPGRILRQTVLD
    EALEPRVAVPQQLITGGVGLALGFWDEEETRSFLVHPETGERLYKTGDLGRVLP
    GNIEFMGREDNOIKLGRYRVELGEIETLKSHPNVRAVIVPVGNDANKLLAVVVP
    EGTERRAAEODASLATERIDARAHAAEADGSLDGRVQFKLARGLRDLGKPVVDL
    TGODPREAGLDVYARRRSVRFTELEAPIFPEVEGRLSCSSVEPDGATLPFERPSAG
    STYPQYTAAYVKSRIEGVDGEGFYVHPFEHRLKLSHGIERGCAHVRONDVDEEA
    PNLLFVGRIDALIESLYGSSRRECLLEGYAAQLMEBAPSCNTGCVQGFNREYOR
    PVLDLRHSDVYVHGMIGRVDPRQVCTGLQDSSPRATTGAPGREQGFADMLRD
    FLKRLPLEYVMPYVFEVLDALPLTNSGVNDRALKALREKDTSSPRHSHTAPDMLLEI
    LVAVYREVLYGLEVYGLQOSFYDLGATSIYRMKSLQKRLDRLEIATILELQYPNLGS
    LASGLRRSRDLDPRNMQDRVEYRRKRRRS"
    16251..21749
    /gene="epob"
    /note="epob"
    /gene="epob"
    /note="EPOS B"
    /codon_start=1
    /transl_table=11
    /product="polyketide synthase"
    /protein_id="AA26920.1"
    /db_xref="GI:6724254"

```

```

Query Match      100.0%   Score 101; DB 1; Length 68750;
Best Local Similarity 100.0%   Pred. No. 8.8e-14;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 AGCGGCTCGAGGCGCTGCTGCCGACGACGAGCGTTCGCGCCGACCGCAGC 60
        |||||||
DB 37682 AGCGGCTCGAGGCGCTGCTGCCGACGACGAGCGTTCGCGCCGACCGCAGC 3741

QY      61 GCGGTGGGTGAGCGAAGCGTCCGATGCTCTGCTCA 101
        |||||||
DB 37742 GCGGTGGGTGAGCGAAGCGTCCGATGCTCTGCTCA 37782

RESULT 3
AR193029 LOCUS AR193029 68750 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 1 from patent US 6346404.
ACCESSION AR193029
VERSION AR193029.1 GI:20238994
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ilgen,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and
Goriach,J.
TITLE Genes for the biosynthesis of epothilones
JOURNAL Patent: US 6346404-A 1 12-FEB-2002;
FEATURES Location/Qualifiers
source 1..68750
/organism="unknown"
BASE COUNT 9596 a 22456 c 25539 g 11159 t
ORIGIN

```


Query Match 100.0%; Score 101; DB 6; Length 68750;
Best Local Similarity 100.0%; Pred. No. 8.8e-14;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGCCGGCTGCGAGGCTGCTCCGACGAGCGTGTCAAGAGCTTCTCGGCCGACCGCAG 60
|||||
Db 37682 AGCCGGCTGCGAGGCTGCTCCGACGAGCGTGTCAAGAGCTTCTCGGCCGACCGCAG 37741
|||||

OY 61 GCGGTGGGTGAGGAGGCTGCGCCATGCTCTGCTCA 101
|||||
Db 37742 GCGGTGGGTGAGGAGGCTGCGCCATGCTCTGCTCA 37782
|||||

RESULT 4
ARI99551
LOCUS ARI99551 68750 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 1 from patent US 6355457.
ACCESSION ARI99551
VERSION ARI99551.1 GI:20249625
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ilyon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and
Gorlach,J.
TITLE Genes for the biosynthesis of epoethilones
JOURNAL Patent: US 6355457-A 1 12-MAR-2002;
FEATURES
source 1..68750
/organism="unknown"

BASE COUNT 9596 a 22456 c 25539 g 11159 t

ORIGIN

Query Match 100.0%; Score 101; DB 6; Length 68750;
Best Local Similarity 100.0%; Pred. No. 8.8e-14;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGCCGGCTGCGAGGCTGCTCCGACGAGCGTGTCAAGAGCTTCTCGGCCGACCGCAG 60
|||||
Db 37682 AGCCGGCTGCGAGGCTGCTCCGACGAGCGTGTCAAGAGCTTCTCGGCCGACCGCAG 37741
|||||

OY 61 GCGGTGGGTGAGGAGGCTGCGCCATGCTCTGCTCA 101
|||||
Db 37742 GCGGTGGGTGAGGAGGCTGCGCCATGCTCTGCTCA 37782
|||||

RESULT 5
ARI99559
LOCUS ARI99559 68750 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 1 from patent US 6355458.
ACCESSION ARI99559
VERSION ARI99559.1 GI:20249633
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ilyon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and
Gorlach,J.
TITLE Genes for the biosynthesis of epoethilones
JOURNAL Patent: US 6355458-A 1 12-MAR-2002;
FEATURES
source 1..68750
/organism="unknown"

BASE COUNT 9596 a 22456 c 25539 g 11159 t

ORIGIN

Query Match 100.0%; Score 101; DB 6; Length 68750;
Best Local Similarity 100.0%; Pred. No. 8.8e-14;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGCCGGCTGCGAGGCTGCTCCGACGAGCGTGTCAAGAGCTTCTCGGCCGACCGCAG 60
|||||
Db 37682 AGCCGGCTGCGAGGCTGCTCCGACGAGCGTGTCAAGAGCTTCTCGGCCGACCGCAG 37741
|||||

OY 61 GCGGTGGGTGAGGAGGCTGCGCCATGCTCTGCTCA 101
|||||
Db 37742 GCGGTGGGTGAGGAGGCTGCGCCATGCTCTGCTCA 37782
|||||

|||||
Db 37682 AGCCGGCTGCGAGGCTGCTCCGACGAGCGTGTCAAGAGCTTCTCGGCCGACCGCAG 37741
|||||

OY 61 GCGGTGGGTGAGGAGGCTGCGCCATGCTCTGCTCA 101
|||||
Db 37742 GCGGTGGGTGAGGAGGCTGCGCCATGCTCTGCTCA 37782
|||||

RESULT 6
ARI99567
LOCUS ARI99567 68750 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 1 from patent US 6355459.
ACCESSION ARI99567
VERSION ARI99567.1 GI:20249641
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ilyon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and
Gorlach,J.
TITLE Genes for the biosynthesis of epoethilones
JOURNAL Patent: US 6355459-A 1 12-MAR-2002;
FEATURES
source 1..68750
/organism="unknown"

BASE COUNT 9596 a 22456 c 25539 g 11159 t

ORIGIN

Query Match 100.0%; Score 101; DB 6; Length 68750;
Best Local Similarity 100.0%; Pred. No. 8.8e-14;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGCCGGCTGCGAGGCTGCTCCGACGAGCGTGTCAAGAGCTTCTCGGCCGACCGCAG 60
|||||
Db 37682 AGCCGGCTGCGAGGCTGCTCCGACGAGCGTGTCAAGAGCTTCTCGGCCGACCGCAG 37741
|||||

OY 61 GCGGTGGGTGAGGAGGCTGCGCCATGCTCTGCTCA 101
|||||
Db 37742 GCGGTGGGTGAGGAGGCTGCGCCATGCTCTGCTCA 37782
|||||

RESULT 7
AR201097
LOCUS AR201097 68750 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 1 from patent US 6358719.
ACCESSION AR201097
VERSION AR201097.1 GI:20251985
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ilyon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and
Gorlach,J.
TITLE Genes for the biosynthesis of epoethilones
JOURNAL Patent: US 6358719-A 1 19-MAR-2002;
FEATURES
source 1..68750
/organism="unknown"

BASE COUNT 9596 a 22456 c 25539 g 11159 t

ORIGIN

Query Match 100.0%; Score 101; DB 6; Length 68750;
Best Local Similarity 100.0%; Pred. No. 8.8e-14;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGCCGGCTGCGAGGCTGCTCCGACGAGCGTGTCAAGAGCTTCTCGGCCGACCGCAG 60
|||||
Db 37682 AGCCGGCTGCGAGGCTGCTCCGACGAGCGTGTCAAGAGCTTCTCGGCCGACCGCAG 37741
|||||

OY 61 GCGGTGGGTGAGGAGGCTGCGCCATGCTCTGCTCA 101
|||||
Db 37742 GCGGTGGGTGAGGAGGCTGCGCCATGCTCTGCTCA 37782
|||||

Db 37742 GGCCTGGGTGAGCGAGGCTGCCCATCTCTCTCTCAA 37782

RESULT 8
LOCUS AR208671 AR208671 68750 bp DNA linear PAT 20-JUN-2002
DEFINITION Sequence 1 from patent US 6383787.
ACCESSION AR208671
VERSION AR208671.1 GI:21509886
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and Gorlach,J.
TITLE Genes for the biosynthesis of epothilones
JOURNAL Patent: US 6383787-A 1 07-MAY-2002;
FEATURES
source Location/Qualifiers
1..68750
/organism="unknown"

BASE COUNT 9596 a 22456 c 25539 g 11159 t
ORIGIN

Query Match 100.0%; Score 101; DB 6; Length 68750;
Best Local Similarity 100.0%; Pred. No. 8.8e-14;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGCGGCTGGAGGCGCTGCCGACGAGCGGTGCAAGAGCTTCTCGCCGACGCCGAC 60
|||||
Db 37682 AGCGGCTGGAGGCGCTGCCGACGAGCGGTGCAAGAGCTTCTCGCCGACGCCGAC 37741
|||||

Qy 61 GGCCTGGGTGAGCGAGGCTGCCCATCTCTCTCTCAA 101
|||||
Db 37742 GGCCTGGGTGAGCGAGGCTGCCCATCTCTCTCTCAA 37782
|||||

RESULT 9
LOCUS AR172664 AR172664 71989 bp DNA linear PAT 17-DEC-2001
DEFINITION Sequence 2 from patent US 6303342.
ACCESSION AR172664
VERSION AR172664.1 GI:17912155
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 71989)
AUTHORS Jullien,B., Katz,L., Khosla,C. and Tang,L.
TITLE Recombinant methods and materials for producing epothilones C and D
JOURNAL Patent: US 6303342-A 2 16-OCT-2001;
FEATURES
source Location/Qualifiers
1..71989
/organism="unknown"

BASE COUNT 10108 a 23531 c 26617 g 11731 t 2 others
ORIGIN

Query Match 100.0%; Score 101; DB 6; Length 71989;
Best Local Similarity 100.0%; Pred. No. 8.7e-14;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGCGGCTGGAGGCGCTGCCGACGAGCGGTGCAAGAGCTTCTCGCCGACGCCGAC 60
|||||
Db 32070 AGCGGCTGGAGGCGCTGCCGACGAGCGGTGCAAGAGCTTCTCGCCGACGCCGAC 32129
|||||

Qy 61 GGCCTGGGTGAGCGAGGCTGCCCATCTCTCTCTCAA 101
|||||
Db 32130 GGCCTGGGTGAGCGAGGCTGCCCATCTCTCTCTCAA 32170
|||||

RESULT 10
LOCUS AB070940/c AB070940 104326 bp DNA linear BCT 22-JAN-2002

DEFINITION Streptomyces avermitilis oligomycin biosynthetic gene cluster.
ACCESSION AB070940
VERSION AB070940.1 GI:15823967
KEYWORDS
SOURCE Streptomyces avermitilis DNA.
ORGANISM Streptomyces avermitilis
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1
AUTHORS Omura,S., Ikeda,H., Ishikawa,J., Hanamoto,A., Takahashi,C.,
Shinohe,M., Takahashi,Y., Horikawa,H., Nakazawa,H., Osone,T.,
Kikuchi,H., Shiba,T., Sakaki,Y. and Hattori,M.
TITLE Genome sequence of an industrial microorganism Streptomyces
avermitilis: deducing the ability of producing secondary
metabolites
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 98 (21), 12215-12220 (2001)
MEDLINE 21477403
REFERENCE 2 (bases 1 to 104326)
AUTHORS Ikeda,H.
TITLE Direct Submission
JOURNAL Submitted (01-SEP-2001) Haruo Ikeda, Kitasato University, Dept. of
Microbial Chemistry, School of Pharmaceutical Sciences, 5-9-1
Shirokane, Minato-ku, Tokyo 108-8641, Japan
(E-mail: ikeda@mc.pharm.kitasato-u.ac.jp, Tel: +81-3-5791-6242,
Fax: +81-3-3444-6197)
FEATURES
source Location/Qualifiers
1..104326
/organism="Streptomyces avermitilis"
/db_xref="taxon:33903"
complement(1..1134)
/gene="mcma"
complement(1..1134)
/gene="mcma"
/note="methylmalonyl-CoA mutase"
/codon_start=1
/transl_table=11
/protein_id="BAB69185.1"
/db_xref="GI:15823968"
/translation="MTGSTADANAREQLVSGTRAPVYFDLPTBAGLSDAPAA
RLVGKGVAVSDIDMRLVFGGLDLDVPSLVADAPAPLVLYOLVAEHNAGAP
RLAGVRGDDVEKEFLARGVGPPLRPSPRLMNAVVAOCALPPOCISVGHILACT
GADPALVFAFTYAGGIEYLRAAVTYAGLVRYAPPLTCLVGGTGGKRRKLRTR
VWTRTANEPFGSGRTAPRLVYTPRDLGLATCTEDTADGATLVDRERGLTLYL
EHGLAQGLGRPHGDRPATPRPHPHQSLDLQAROTERLAKLRMRVQPAVDALLR
LRKAAGDNDVLYPMREALLAATVGEVCDALRSTWGTGPSAT"
complement(1131..1886)
/gene="ppia"
complement(1131..1886)
/gene="ppta"
/codon_start=1
/transl_table=11
/product="phosphopantetheinyl transferase"
/protein_id="BAB69186.1"
/db_xref="GI:15823969"
/translation="WRAHEVFADVPVSLHPREALVODAVESRRERFATGRHCARRA
LASLGVPPGPLATPRGAPRWPAGVCGAITHCAGRAAARAPGIRAVGIDAEPRP
LPEGVLEAVLPAEOFOYKEHTRARHPGVWMLFSAKESYKAFPGATGRLGEERA
LIDLPVASTFERYRLPTAIRAAREPDLIDLGRLMYRHGIIIVTVAVVVGPPDRDPNAP
SPYDRTRKSAPTRGSADRRGCRPTDRATDRPTA"
complement(1910..2875)
/gene="simx4 homolog"
complement(1910..2875)
/gene="simx4 homolog"
complement(1910..2875)
/gene="simx4 homolog"
/codon_start=1
/transl_table=11
/product="simx4 homolog"
/protein_id="BAB69187.1"
/db_xref="GI:15823970"
/translation="MSDRHDSSTALGSPSHGRRLAVSDIHYRHRRNRDLYVDLHP
GCAEDMLIAGDVAVESVADITWALSILDRRAKYVMAGNHEIMTAPRQDDELGERK
YRHLVDICRALGVYTPEDPYFVWNGAGCPVAVAFELIYDYSFLPRGMSKSHLRILA
DDAGVTCDFEFLHDPYPSRAWCRAVRAVATRAIDLPLDPLVLVNHPLVREPT
RYLRHPEFALWCCTTATADWHLRYRAAIVVGHILHPIPTIRCDVPHHEVSLGYPREV

ORYGRGRPMWPVOILLTPAPAPAPAPVPPVPAPRPPCHRTASPADRRP"
3213.3971
/gene="OlmC"
3213.3971
/gene="OlmC"
/codon_start=1
/transl_table=11
/product="thioesterase"
/protein_id="BAB6918.1"
/db_xref="GI:15823971"
/transl_table="MSANRVDGNLVMROFHPAPPAATKLVCLPHAGGSAHYFLPVSKA
LAPGVDAIOYQPGRODRHHEKCIDNTIAELADATVATLEPADPITLPHSGMCAITLA
FEVALRLEKTIYPLALFASGRRAAPSNHRDESVHLRQDDEIITAEIKTLSTSTRSEYLD
EELLMLPRLASDRTKAETRYVAGACATLAPHAHGVDEDPVSLDEADQAMADHTTG
HEELMSHPGHPHYLNDQAPVIEIARVSSVAS"
4643.7438
/gene="OlmRI"
4643.7438
/gene="OlmRI"
/codon_start=1
/transl_table=11
/product="luxR-family transcriptional regulatory protein"
/protein_id="BAB69189.1"
/db_xref="GI:15823972"
/translation="MLIGRGEIDRLRANSSVNSTGRSAVYESAGCGKSEILD
LVRAHYESGLIALRATATAEERNTPLSYLRSLVNGAPFPDLYRFFTHMSDAAARMS
TEGRKASGADLAAWELMRTCTEVRALAQARPILVVAQDIDHIGDADSLRLHLIARIS
RNVPLATLFTFQSLAGPADPFGAEILRQPLMEIRPRECRDIOAMLEBOGRADG
OGIVRYRVAAGSNPDLRALLEDRALEPDPTVADQVPAGQAVGRVATACIDRGG
EKAHVAGSLAVIGGAPFHELISRLDLSAALTENIRALRAAGIYVGTSPORAAKA
AVDSLEPFSHLRHOSSAULLHSGAPAAITAAEHLDEAGQADGAMAVGYLOQAAEQA
LLADEARAVSYLELAHAASADPPRGTEIKIRIGSVTRVSAADAEHRVDALDFEQA
GLPLSPDGGIQAALLAHAGLEQARSVLEQGRPGDDADRTARQIAETRDSYLNR
GAWCGTGAAMHENDPAGRPSTYRAGTGLRAWLSGSHALLAORIACTEGLEITLHLD
ATFDVLYAVLTLAIGAVERALHMCOTFLEQVQADQAPGMEVFAVRABGALRGSN
LAELARDARRCFPAVLPERRRSALFEGSAVSYIVYQIRAKGTEADAAROLSRPADSLPD
TTHIMDIYRAGGLTYMATRQYHSALRDBHEHGRLAEMKGTDMPLWPKWTDTEALLR
LGEROQIGRLATEQULAVQDGAAPRIKGISLQALQAPDILNRAVEYSRADDE
LGERSTRTLEDADHRETRQOAMVYLRARQOLKSCGAPPPDRIGSPPDAGASGE
LFGEDPENNAADMGVLESERERVAGIASGYNREISVRLYITNSTVEQHLTRYRKAK
INGRELLAEARAGVQV"
7658.10486
/gene="OlmRII"
7658.10486
/gene="OlmRII"
/codon_start=1
/transl_table=11
/product="luxR-family transcriptional regulatory protein"
/protein_id="BAB69190.1"
/db_xref="GI:15823973"
/translation="MLASAGOGEGGVVLYNAGPEVGRITALIDAFILRKATESGAVCA
AGSSAEVTAEEGVVGOLEPPTGPIGAGRLARLADSGADGIDADLPAILOEILRR
RPTVAYVNDQADLAASAISRLRYLRLRTPA PVMWVLTPEPDPALPSPVOALRRP
OCRRIGILPLSRDEYTOVYSKQVSDRYPLADPFHISGNNPLLYTALLADHRAARK
LSPVAFVTFVSSVSSAASASASASVSAAPASAAFAAALAMVYREVPYLE
DVACFVALGESAIPARVACLVDRGADAVDVMYALDTAGIAGGTFRSALVGAALLE
HYDVTAREILRRRARLHLDGAPATVVARLILAAPAEPPAGITFLAAAEKAMQESR
BELSIDCRLRAGOSATTEOBERAAVVMARVKIWEIDRPLVSPWDELVVALRQDLS
GDAIVTYLHMDORPEPAVRSISTIAAGADGAPGVYAEILHYRHLHSRCSPTLAS
VPGSNPIPRPTPLRLRRVSPASTYAGEVLGKVTGEGPDHVTYAAEBETLIRKCRSETYV
EALAEALIALATVYAGCPDAAAFWCGDGLQOAGTRSTGTSVALLSGIRAEIALRQDALPA
VEOYKRSALISAISLPGMGVNAAGSPILAAVLHSTTASGMSGVAGMWLNQAVDPGMFTRH
GLIYTHAGHGHILADPVAALIDDFITCGGLAKEMGMDVPTFLPRTSAQAOALQALN
PGQAAVIRBEOLDRGGSTTARAGVALRLIATNAGDEGRPALLESVNLIESGSDPGV
LIRALADBCOLDHYGAPAKAKARMTARRAHRAAADCCGADLPFRKIFREVPYDRBEGDPC
LTGEOGGSASLTAAEBRRVTALAAIGHSKREIIGRKLFITKSTVEOHLTRYRKLGVRNR
ADLGDLAGTINLGAOTGPAMMASSAVAG"
10933.11271
/codon_start=1
/transl_table=11
/product="p450-like protein"
/protein_id="BAB69191.1"
/db_xref="GI:15823974"

gene 29989
/gene="o1m1"
11458, .29989
CDS
/gene="o1m1"
/codon_start=1
/strand=+
/product="modular polyketide synthase"
/protein_id="BAB69192.1"
/db_xref="GI:15823975"

translocation="MACRIPGAAGPAEFLALLRNGEDAVGAAPDRRYARPRGFLDS
VVEEDGFVFGISPREAAMDPOQRLVLELMEALDEGASIPARLGGSSQSSLVFVAILD
DVATLISRAAGADAAAPOTTGTGNGIILNKRYSYTLGGSPFTDQSSQSSLVFVAILA
ALAEIRGEGSVLAAAGVNLILAEIDSTVAVERFGALSPGYCYTDAANGVRRGEGG
VYVLRKADAIADGGDIHCLILAGSVNNDDGEGELTVYDRRAGQVLAATVAEAGISIP
LAKVAIVELHGTGPTGDPVEAAAVGAVIACARPGOPILNLSVNTNGHLEGAAGTIG
LKAVALVAHREIRPATLHHASPNRIPPAELGICVNTALITPMPQDAPRLVYGVSSBOM
GGNCHVILVTEPAGTIDAAARPATGIRPAGTIPAPLIVSGRDEQALNQAQRRLRDLDT
HPLDLPDVAHTLAAARAPLARAVALLAAREDEASIGLRALADTPOADIVLFAAGVSG
ALAFVTEGOSORPGMTAEILRTSAEYAAALDEVCASIDPLRLAPVTEFVFAAGSBE
AALDTEFTPOLALFPAVEALFEYRAEHGDLFPRLTGHSEVLEAAHVAAGLSTADAC
TIYAAAGTRMODRPATGAVNSVQATEDEGLAPRFSLSYTLAENLPASTYVAGCEDNLA
VAHAHAARGRKTRLRVNSHAFSHMGOMLTERRIIAGLTFCAPPIPIVSNETGTL
LEADRSRPEYVWRHARAARVRLFDGVLLEDDQVYTTLELPGDVLISAMARDCLRQD
TCVPLRLGRDREYTEREALLAALAGLVYGVGEMMTERTTEGARVRVLPYTAQRHRH
WLEPESASRPVNGGAGAGGAGAGGAGAGGAGAGGAGAGGAGAGGAGAGGAGAGGAG
DRTTLELVBDATAOYLIGHASPDVATHEAPRLEFGPSSLMIALEGRLERTVATGRVPT
LLEPHTPOLAIRLEADGAERTDAGPSPAPRAAHDDPILVITYMGCRIPGVGRSPE
DLMRPFAAGGDAISALPVDNGWFERADPREGGLADYAEHEDAEFTGTSIPREALAMP
OQRLLETTEWEALERAGVDALKLRGSRVGVNAGSPSEYGRHETSDAGHVALTGA
PSYLSGRIRIAYVLGRLBPALVDTJACSSILVLIHAAQALRGEBDDLAAAGVYVMTS
GMAEAFRGGGLAPDRCKAFADGADGTGMEGVGMVLAORLSAVIDRGPVLAIVNG
SANSNGASNGLTPANPGPSOORVTRQALASAGLSADVDVEAHGTGSLGDPTEAO
LLATYQDPRDPRPILMGLSVKSNIGHTQAAAGVAGVKSVALRNGCILTPTIHYDEPT
PEVMSAGAVELLTBERWEPETDGERRRGVSATFGISGTNAHYILEQAPVSEEPYTA
DDGLVPMVVASRSEELRAQRLADHVRBSGLSPADVAGLSLAVTRAGLEHRAVLVA
DEBAGFLAHLDALABGRTPAGVYVEGCTAFLFSGGSGORACMGRELYSAHPVAAAD
EVCAAFDGLLPGALREVYVAGGLDRTGMAQDPALEFALFEILVSSMGVADVLLGH
SIGELAAIVAGVWSLADACRIVARGLSMOALPFGGAMVVEAAEBELPDLPGGVY
AANNGRSVLISGDEPYPYALQAFABDGRRTKIOYSHAFSARMMPMLAEFEETLA
VAEFRRPIPVVSNVTEMAWGEFTTPIPVNREHREAVRPLDGTGCAAHAGVYRETEL
GPGDTLTJLHADHDAGDVAVALPKGDIDPERTVTVAAGRLYATGLVADWPAFA
PYGARTELEPTAFQHRRYMTEPVAPSSAPPANLRYRTAAVASPDGPRSLSGTNAV
VPGASSEDIYAGVYTAALRDHGAPRVYVEDPAGPEADVAOTLRALGEDAGVVS
LGTGASTSSGVAATAVIVYQAWTIDVAGSDGAPLWCLTRGASVQDPAQOALMGM
GRRAALEHPRMNGCLVDLPAYVEDDPAKTRILCAIVLAGSGSEQVAVYRAAGLFAKLGA
GRADAPAEAPSPWRTDGTIVLTGGTGALGHAVLQWLAAAGAAHVLLTISRGPRADGA
AETLTALIRETGETEYVAACDVTDRLALIELSGLRPDLPLGVYHAAVGDVDDGLSL
TTPRFAAVAGPKVTGAMHLHELTRDLDAELSSFEFGYGLGAQNOYAAANVYLDLA
LAWHAROLDPATPAVAMGMSNGSAGMAATPEAAROLDATGTGVLPPARAPALALGQVND
DGETATVADVDMDRRAAGFGGRSPSLITGVPEVRRLEERHETGRACAPAEVTRDLSL

Query Match 59.2%; Score 59.8; DB 1; Length 104326;
Best Local Similarity 76.8%; Pred. No. 0.0016;
Matches 73; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

4 CGGCTGGAGGCGCTGCTCCCGAGAGCGGTGCAAGCTTCTCGGCGCCGACGAGCGC 63
DB 100340 CGGAGCGGCGCTCGGCCCGAGCGAGGAGGTAAGTCTTCGCGCGCCGACGCT 100281
Oy 64 GTGGGCTGAGAGCGAAGCGCTCGCCATGCTCTCTCT 98
DB 100280 GTCCGGCTGGGCGAGGCGCTCCGCGCTCTCTGCTGT 100246

RESULT 11
AF411573S4 3185 bp DNA linear BCT 27-SEP-2001
LOCUS Actinomadura verucosospora polyketide synthase 2 gene, complete
cds.
AF411576

VERSION AF411576.1 GI:15788318
 KEYWORDS 4 of 4
 SEGMENT
 SOURCE
 ORGANISM
 Actinomyces verrucosuspora.
 Actinomyces verrucosuspora
 Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 Actinomycetales; Streptosporangineae; Thermomonosporaceae;
 Actinomyces.
 REFERENCE 1 (bases 1 to 3185)
 AUTHORS Tsantrizos, Y.S. and Yang, X.
 TITLE Cloning and characterization of a gene cluster involved in the
 biosynthesis of verucopeptin, a cyclodeipeptide metabolite of
 Actinomyces verrucosuspora
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 3185)
 AUTHORS Tsantrizos, Y.S. and Yang, X.
 TITLE Direct Submission
 JOURNAL Submitted (19-AUG-2001) Department of Chemistry, Boehringer
 Ingelheim (Canada), Laval, Quebec H7S 2G5, Canada
 FEATURES
 source
 1. .3185
 /organism="Actinomyces verrucosuspora"
 /strain="0886-2"
 /db_xref="taxon:46165"
 CDS
 1. .3126
 /codon_start=1
 /transl_table=1
 /product="polyketide synthase 2"
 /protein_id="AF107759.1"
 /db_xref="GI:15788322"
 /translation="MDDADAPADGADVERALRLLEEEDRLRRITDILKAGREGE
 IAAVAMACRPFGVSTPEGLAEAVRGGDIVEEPKRWMDVDEDEGTLGTSSTR
 GGFVTVAGDFAEFFGSIAPREALAMPDQRLLEASWEFERAGIVPDVGTGAGV
 SGVSPGXGLFLEVDHLEHGLSLGSLASVAGRAVYQLGTPAMTVTACSSSI
 VALHIAVRLSRGECSLAGGAAMVTPAYFEFRSMOROGAPGRCAAFADADG
 MSEVGVLLERLISDADNRGHPVLAIRGSAVNDGASNGITASGRQEEVITKALA
 DAGTAASEVDLVEAHGTGTVDITLHALMTTGGQAEADRLPLMSLKSNIHQAO
 AAGVAGYKAAVMAIRHGILPKTLHLDAIPSRVDMSCGAVELLVGRERMDGPRAG
 VSSGFSGTNAHVILIEAPRKGAALAEPIPGPVALGGLVPAVSGKNPAAALRQAALR
 DFVTVSYGAQADADIDGMSLVSRAFAHRAVVLGRDDELISGLASEGAEAGVV
 RGVGELGVGVFMPGGGOWAGMGRRLYDAYPYFAHALDECAALDWMAGMSLLDYV
 RGEAGAPLEIVDVYOPALBSVYTMALMSKVEPRAAVYVHSGEITAAHVCGALSL
 LRDAAVVALSRALVDIADHGGMAVSAESDAVAEOLAPSRHLGAAVGPSSVAV
 SGALADALDEFLEKNNAGVARRISVYASISHEVARREVIPLSVNGRTSELEPE
 YSTLYGEVIDIAELDDGYMYNLRKYLFESSVRLADDGVRFTIEMSPHVLTPVO
 EMVADGCAVYLVSSRRDRGEYEAAMDSLAOLYVGGVYDMAALGTRRRITLPYARG
 RRRWLNSTHVAVMELPADAPAEEDQVPLPRVVALPBGDAEALYDHLVETKMT
 VLGAPSESDVDPDEFKIDGLFDSLADLSRLAASIGVRLKPTLVLRHPSRLIAGH
 IAAVMDRRRA"
 BASE COUNT 395 a 1086 c 1254 g 450 t
 ORIGIN
 Query Match 58.6%; Score 59.2; DB 1; Length 3185;
 Best Local Similarity 76.0%; Pred. No. 0.00047;
 Matches 73; Conservative 0; Mismatches 23; Indels 0; Gaps 0;
 QY 3 CCGGTCGAGAGCCCTGCTCCGAGAGAGGTCGAAGAGCTTCTGGCGCGAGCGAGG 62
 Db 762 CCGGAGAGCGGCGCTCGGCGCCGAGAGCGCGGTGCAAGCGCTTCGCGGAGCGCGCGAGCG 821
 QY 63 CGTGGGTGAGCGAAGGCTCGGCCATCTCTCTCT 98
 Db 822 GACCGGCTGTCTCGAGGCGCTCGGCTCTCTCTCT 857
 RESULT 12
 LOCUS AF016585 4109 bp DNA linear BCT 07-DEC-1997
 DEFINITION Streptomyces caelestis cyclochrome P-450 hydroxylase homolog (nidi)
 genes, partial cds; polyketide synthase modules 1 through 7 (nidi)
 genes, complete cds; and N-methyltransferase homolog gene, partial
 cds.
 ACCESSION AF016585

[illegible]

[illegible]

RESULT 7	LOCUS	DEFINITION	ACCESION
AQ797032	437 bp	DNA linear	GSS 04-AUG-1999
nbxb0071IN03r	CUGI Rice BAC Library	Oryza sativa genomic clone	
nbxb0071IN03r	DNA sequence.		
10207022			

ACCESSION	AQ/9/052
VERSION	AQ97032.1
KEYWORDS	GSS.
SOURCE	Oryza sativa.
ORGANISM	Oryza sativa

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
1 (bases 1 to 437)
Wing, R.A. and Dean, R.A.
A BAC End Sequencing Framework to Sequence the Rice genome
Unpublished (1998)
Contact: Wing RA

Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: r4rld@clermson.edu
Seq primer: GCGAACAAGCTATGACCATG
Class: BAC ends
High quality sequence stop: 233.

FEATURES	Location/Qualifiers
source	1. .437

```

54 a      148 c      149 g      85 t      1 others
BASE COUNT
ORIGIN
/oranism="Oryza sativa"
/strain="Japonica"
/cultivar="Nipponbare"
/db_xref="taxon:4530"
/clone="ndxb0071N03r"
/clone_lib="CGI Rice BAC Library"
/tissue_type="Leaf"
/lab_host="E. coli DH10B"
/notes="Vector: pBelBAC11; Site_1: HindIII; Site_2:
HindIII; Rice is one of two most popular grains in the
world. Half of the world population especially those
inhabiting highly populated areas of the humid tropics
and subtropics, rely on rice as their primary source of
carbohydrate. Monocotyledonous rice is a diploid plant
(2n=24) with a haploid genome equivalent of 431 Mbp
(Arumanathan and Earle, 1991). The relatively small
genome of rice, three times larger than that of
Arabidopsis, makes it suitable for genomic studies. In
order to facilitate positional cloning, physical mapping
and genome sequencing of rice, we have constructed a BAC
library from Oryza sativa, Nipponbare variety. The
library contains 36,864 clones with an average insert size
of 128.5 kb providing 10.9 haploid genome equivalents. The
deep coverage allows the isolation a particular sequence
with a probability of 99.9 %. Two high density filters,
each containing 18,432 clones (doubly spotted), represent
the whole library for colony screening."
```

Query Match	33.1%	Score 33.4	DB 17	Length 437
Best Local Similarity	65.3%	Pred. NO. 71		
Matches 49	Conservative 0	Mismatches 26	Indels 0	Gaps 0
07	15	CGCGGCGGTTCATGAGCGCGCGCGCCCGTCGATCGGCGCGCTGTGGCA	74	

Db	Qy
278	75
CGTGGCGCTCCGCTCCGAGGGGTTGCGAGCTCGGTGCGCCCTGCACCCGCGCGCTGTTCCG	CTTGGGCTCGACTC
337	89
CGTGGCGCTCCGCTCCGAGGGGTTGCGAGCTCGGTGCGCCCTGCACCCGCGCGCTGTTCCG	CTTGGGCTCGACTC
338	352
CGTGGCGCTCCGCTCCGAGGGGTTGCGAGCTCGGTGCGCCCTGCACCCGCGCGCTGTTCCG	CTTGGGCTCGACTC

RESULT 8	LOCUS	DEFINITION
AV628827	384 bp	linear EST 15-DEC-2000
AV628827	Chlamydomonas reinhardtii 5% to 0.04% CO2 Chlamydomonas reinhardtii cDNA clone LC1047f08_r 5', mRNA sequence.	

ACCESSION	AV62882/	
VERSION	AV628827.1	GI:10791461
KEYWORDS	EST.	
SOURCE	Chlamydomonas reinhardtii.	
ORGANISM	Chlamydomonas reinhardtii	

REFERENCE
AUTHORS
TITLE
JOURNAL

Eukaryota: Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales
Chlamydomonadales; Chlamydomonas.
1 (bases 1 to 384)
Asamitsu, E., Miura, K., Kuchio, K., Inoue, Y., Fukuzawa, H., Ohyama, K.,
Nakamura, Y. and Tabata, S.
Generation of expressed sequence tags from low-CO2 and high-CO2
adapted cells of Chlamydomonas reinhardtii
DNA Res. 7 (5), 305-307 (2000)

FEATURES	COMMENT	MEDLINE
20539644	Contact: Erika Asamizu The First Laboratory for Plant Gene Research Kazusa DNA Research Institute Yana 1532-3, Kisarazu, Chiba 292-0812, Japan Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/Location/Qualifiers	

```

source
1. .384
/organism="Chlamydomonas reinhardtii"
/strain="C9"
/db_xref="taxon:3055"
/clone="LC1047f08_r"
/clone_lib="Chlamydomonas reinhardtii 5% to 0.04% CO2"
/note="Vector: pluescriptII SK-; Site_1: EcoRI; Site_2:
XhoI; The cDNA library was constructed from cells cultured
in a carbon stress acclimatized condition in which carbon
dioxide concentration in the bubbling gas was changed from
5% to 0.04%"
56 a 122 c 160 g 46 t
BASE COUNT
ORIGIN

```

Query Match	32.9%	Score 33.2;	DB 10;	Length 384;
Best Local Similarity	64.1%;	Pred. No. 78;		
Matches 50;	Conservative	0;	Mismatches 28;	Indels 0;
				Gaps 0

Qy 2 TGCAGCCGAGATCGCCGGCGTGTCTTATATGGGGGCCGACGGCGTTGCCGTCATC 61
 ||||| | | | | | | | | | | | | | | | | | |
Dd 262 TGCAGCCCAAGTTTCACAATGCACTGGCTTGCTGCGACGCGCCGACACTCAAGC 322

QY 62 GCGCGCTGTCGACTTG 79
||| ||||| |||
Db 322 AGCTGAGTCGAGCTGG 339

RESULT	9
AV432901	
LOCUS	502 bp mRNA linear EST 23-AUG-2000
DEFINITION	AVA32901 Porphyra yezoensis TU-1 Porphyra yezoensis cDNA clone
ACCESSION	PW021e05_1 ⁻ 5', mRNA sequence.
VERSION	AVA32901.1 GI:8588126

KEYWORDS EST.
SOURCE Porphyra yezoensis.
ORGANISM Porphyra yezoensis.
Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae; Porphyra.
REFERENCE 1 (bases 1 to 502)
Nikaido, I., Asamizu, E., Nakajima, M., Nakamura, Y., Saga, N. and

LOCUS	DEFINITION	AV924405	542 bp	mrna	linear	EST 18-JAN-2002
AV924405	AV924405 K. Sato unpublished cDNA library, cv. Haruna Nijo second leaf stage seedling leaves	Hordeum vulgare subsp. vulgare				
AV924405	clone basd14g21 5', mRNA sequence.	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae				
AV924405	AV924405					
AV924405	AV924405.1 GI:18220184					
AV924405	EST.					
AV924405	Hordeum vulgare subsp. vulgare.					
AV924405	Hordeum vulgare subsp. vulgare					
AV924405	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae					
AV924405	Trilliceae; Hordeum.					
AV924405	1 (bases 1 to 542)					
AV924405	Sato, K., Salscho, D. and Takeda, K.					
AV924405	Bailey EST sequencing project in NIG and Okayama Univ					
AV924405	unpublished (2002)					
AV924405	Contact: Tadasu Shin-I					
AV924405	National Institute of Genetics					
AV924405	1111 yata, Mishima, Shizuoka 411-8540, Japan					
AV924405	Tel: 81-559-81-6856					
AV924405	Fax: 81-559-81-6855					
AV924405	Email: tshini@genes.nig.ac.jp.					
AV924405	Location/Qualifiers					
AV924405	1..542					
AV924405	/organism="Hordeum vulgare subsp. vulgare"					
AV924405	/cultivar="Haruna Nijo"					
AV924405	/db_xref="taxon:112509"					
AV924405	/clone="basd14g21"					
AV924405	Nijo second leaf stage seedling leaves"					
AV924405	/tissue_type="seedling leaves"					
AV924405	/dev_stage="second leaf stage"					
AV924405	90 a 200 c 169 g 83 t					
AV924405	BASE COUNT					
AV924405	ORIGIN					
AV924405	Query Match					
AV924405	Best Local Similarity 32.7%; Score 33; DB 10; Length 542;					
AV924405	Matches 57; Conservative 0; Mismatches 40; Indels 0; Gaps 0;					
AV924405	1 GTGACGCGCGAGATCGCGCGCTCTTCATGAGGCGCGCGAGCGCGCTGCGTCAAT 60					
AV924405	111 111 111 111 111 111 111 111 111 111 111 111 111 111 111					
AV924405	404 GTGTCGGCGAGCGGCGCGCGCGGTGATCTGATGAGCGCGCGACCGCTGAGCGCTCGAG 463					
AV924405	61 CGGCGCGTGTGACCTGTGAGCGCTGACATCGCGTCAAGG 97					
AV924405	111111 111111 111111 111111 111111 111111 111111 111111					
AV924405	464 CGCCGCGCTTCGAGATGATGTCGCGCTCAGACAGCCG 500					
AV924405	RESULT 13					
AV924405	LOCUS					
AV924405	AL818682					
AV924405	DEFINITION					
AV924405	AL818682 1:125 Trillium aestivum cDNA clone B06_1125_plate_12, mRNA					
AV924405	sequence.					
AV924405	AL818682					
AV924405	AL818682.1 GI:21830282					
AV924405	EST.					
AV924405	bread wheat.					
AV924405	Trillium aestivum					
AV924405	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;					
AV924405	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae					
AV924405	; Trilliceae; Trillium.					
AV924405	1 (bases 1 to 513)					
AV924405	Wilson, I., Beswick, R., Shepherd, S., Barker, G., Parker, J., Owen, P.,					
AV924405	Edwards, D., Cognill, J., Holdsworth, M., Lenton, J., Shewry, P. and					
AV924405	Edwards, K.					
AV924405	A BSRN-funded wheat EST resource for the academic community					
AV924405	unpublished (2002)					

<hr/>					
Institute of Arable Crop Research					
Long Ashton, Bristol BS41 9AF United Kingdom.					
<hr/>					
FEATURES	Location/Qualifiers				
Source	1..513 /organism="Triticum aestivum" /cultivar="mercia" /db_xref="taxon:4565" /clone="B06_1125_plate_12" /clone_1lb="1:125" /tissue_type="embryo" /dev_stage="30 days post anthesis"				
BASE COUNT	103 a 135 c 172 g 102 t 1 others				
ORIGIN					
<hr/>					
Query Match	32.5%, Score 32.8; DB 9; Length 513;				
Best Local Similarity	59.8%; Pred. No. 1e+02;				
Matches	55; Conservative 0; Mismatches 37; Indels 0; Gaps 0;				
OY	1 GGCGAGCCGGAATGGCGCGTGCTTCATNGGGGCCGCAGCGCCGTCCCTCAT 60				
DB	185 GTGGCGGCTCGGGGGCGCGCGCCTGGCTTCAGGGACGCGCGAGGCCCGTGGCGGGTACA 244				
OY	61 CGGCGCGTGTGCGACTTGAGCCCTCACATCGCT 92				
DB	245 ACGCGCGCGCGCGCGGTGGCAGCGCAACCAAGCT 276				
<hr/>					
RESULT 14	1565 bp mRNA linear EST 12-MAR-2002				
BM909387					
LOCUS	AGENCOURT.6640954 NIH_MGC_99 Homo sapiens CDNA clone IMAGE:5433958				
DEFINITION	5', mRNA sequence.				
ACCESSION	BM909387				
VERSION	BM909387.1 GI:19359766				
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 1565) NIH-MGC http://mgc.nci.nih.gov/. National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)				
AUTHORS	Contact: Robert Strausberg, Ph.D.				
TITLE	Email: cgapbs@email.nih.gov				
JOURNAL	Tissue Procurement: Lou Straud				
COMMENT	CDNA Library Preparation: Rubin Laboratory CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LML) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LML at: http://image.lml.gov Plate: LICM1910 row: d column: 23 High quality sequence start: 32 High quality sequence stop: 137.				
FEATURES	Location/Qualifiers				
Source	1..1565 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:5433958" /clone_1lb="NIH_MGC_99" /tissue_type="lymphoma, cell line" /lab_host="DH10B (phage-resistant)" /note="Organ: lymph; Vector: pOTB7; Site:1: XhoI; Site:2: EcoRI; CDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCGAGAG(g). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-CDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC library."				
BASE COUNT	73 a 696 c 376 g 401 t 19 others				

ORIGIN

Query Match 32.5%; Score 32.8; DB 14; Length 1565;
Best Local Similarity 59.8%; Pred. No. 1.4e+02;
Matches 52; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

QY 15 GCGCGCGCTCTTCATGAGGCGCGGAGCGCGCTCCGTCGATCGCGCGCTGTGGA 74
||||| ||| | ||||| ||||| ||| ||| ||| |||
Db 1259 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTGTGCGTGTGCGCGCGCTGTG 1318
||||| ||| | ||||| ||||| ||| ||| ||| |||
QY 75 CTGGGCGCTGACGCTGCTACGCGCGT 101
||| ||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1319 CCTGCGCTGNCGCTTCCGCTCCGCGCT 1345

RESULT 15

BE859603

LOCUS BE859603 345 bp mRNA linear EST 29-SEP-2000
DEFINITION UI-M-AP0-abe-c-05-0-UI.r1 NIH_BMAP_MST Mus musculus cDNA clone

ACCESSION BE859603

VERSION BE859603.1 GI:103755692

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 345)

AUTHORS Donald, M.F., Lennon, G. and Soares, M.B.

TITLE Normalization and subtraction: two approaches to facilitate gene
discovery

JOURNAL Genome Res. 6 (9), 791-806 (1996)

MEDLINE

COMMENT

Contact: Chln, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
20892-9643, USA
Tel: 301 443 1706
Fax: 301 443 9890
Email: MEST@mail.nih.gov
CDNA Library Preparation: M.B. Soares Lab Clone distribution:
Researchers may obtain BMAP cDNA clones from RESEARCH GENETICS. It
should be noted that Bento Soares is generating a small number of
additional specialized non-redundant arrays of BMAP cDNAs whose
availability will be considered under appropriate and limited
collaborative arrangements
Seq primer: M13 Reverse

FEATURES

source

Location/Qualifiers

1..345

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="UI-M-AP0-abe-c-05-0-UI"

/clone_lib="NIH_BMAP_MST"

/dev_stage="27-32 days"

/lab_host="DH10B (Life Technologies)"

/note="Vector: p773D-Pac (Pharmacia) with a modified
polylinker. Site 1: Not I; Site 2: Eco RI; The
NIH_BMAP_MST library is a non-normalized library
constructed from mouse scrotum. The tag is a string of 5
nucleotides present between the Not I site and the
oligo-dT track. The library was constructed as described
by Donald, Lennon and Soares, Genome Research 6: 791-806
, 1996. Tissue provided by Ms. Anne Novakovich,
Zivic-Miller Laboratories."

BASE COUNT

49 a 125 c 123 g 48 t

ORIGIN

Query Match 32.3%; Score 32.6; DB 12; Length 345;
Best Local Similarity 58.9%; Pred. No. 1.1e+02;

Matches 56; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 6 GCGCGAGATCGCGCGCTTCATGAGGCGCGGAGCGCGCTGCTGATCGCGCC 65

Db 1 GCGGAGGACGACGAGCCTCGGCTTCCCTGGCGGAGAGCCATGATGTCGCGCGGCT 60
||||| ||| ||| ||||| | ||||| ||| ||| ||| |||

QY 66 GCTGTGAGCTTGAGGCTCGACTGCTGACGCGGCGG 100
||||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Db 61 GCTGGCGGCTGCTGCTGACACCGGCGGCTGTGG 95
||||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Search completed: November 6, 2002, 15:52:27
Job time: 1202.62 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 6, 2002, 13:32:49 ; Search time 27.125 Seconds
(Without alignments)
1240.503 Million cell updates/sec

Title: US-09-724-876-2_COPY_31170_31270
Perfect score: 101
Sequence: 1 gtgcagcgccgagatcgccg.....ctgcactgcctcagcgcggt 101

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 310279 seqs, 166577418 residues

Total number of hits satisfying chosen parameters: 620558

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Published Applications_NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
- 10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
- 12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	38.6	38.2	4041	10	US-09-861-289-36
2	38.6	38.2	36778	10	US-09-861-289-5
3	33.8	33.5	11220	10	US-09-861-289-32
4	33.8	33.5	13842	10	US-09-861-289-30
5	30.6	30.3	15872	10	US-09-861-289-1
6	30.4	30.1	3189	10	US-09-815-242-4056
7	29.6	29.3	3209	10	US-09-925-301-474
8	29.4	29.1	99	10	US-09-815-242-1227
9	29.4	29.1	1413	10	US-09-815-242-7980
10	29	28.7	1274	10	US-09-919-172-93
11	28	27.7	1140	10	US-09-861-289-15
12	28	27.7	1326	10	US-09-749-7288-12
13	28	27.7	13613	10	US-09-861-289-3
14	27.6	27.3	1320	9	US-09-967-4778-3
15	27.4	27.1	446	10	US-09-960-352-8797
16	27.4	27.1	473	10	US-09-960-352-1195
17	27.4	27.1	1197	10	US-09-960-352-1195
18	27.4	27.1	1197	10	US-09-796-338A-18
19	27.4	27.1	1609	10	US-09-897-201-1
					Sequence 16, Appl

20	27.4	27.1	2700	10	US-09-826-508-25	Sequence 25, Appl
21	27.4	27.1	2823	9	US-09-793-139-3	Sequence 3, Appl
22	27.4	27.1	2823	10	US-09-818-879-3	Sequence 3, Appl
23	27.4	27.1	2823	10	US-09-211-755B-3	Sequence 3, Appl
24	27.4	27.1	2826	9	US-09-793-139-46	Sequence 46, Appl
25	27.4	27.1	2826	10	US-09-818-879-46	Sequence 46, Appl
26	27.4	27.1	2826	10	US-09-211-755B-46	Sequence 50, Appl
27	27.4	27.1	12425	12	US-10-023-529-50	Sequence 50, Appl
28	27.4	27.1	12425	12	US-10-023-523-50	Sequence 50, Appl
29	27.2	26.9	3345	10	US-09-894-998-49	Sequence 49, Appl
30	27	26.7	822	10	US-09-815-242-9914	Sequence 914, Ap
31	27	26.7	2998	10	US-09-752-639-4	Sequence 4, Appl
32	27	26.7	2998	10	US-09-752-639-149	Sequence 149, App
33	27	26.7	2998	10	US-09-984-198-4	Sequence 4, Appl
34	27	26.7	2998	10	US-09-984-198-149	Sequence 149, App
35	26.8	26.5	666	10	US-09-780-717-6	Sequence 6, Appl
36	26.8	26.5	981	10	US-09-780-717-4	Sequence 4, Appl
37	26.8	26.5	1163	10	US-09-736-131-1	Sequence 1, Appl
38	26.8	26.5	1170	10	US-09-823-829-13	Sequence 13, Appl
39	26.8	26.5	1510	12	US-10-052-586-589	Sequence 589, App
40	26.8	26.5	12380	10	US-09-736-131-3	Sequence 3, Appl
41	26.6	26.3	1701	10	US-09-933-814-1	Sequence 1, Appl
42	26.6	26.3	1701	10	US-09-824-134-1	Sequence 1, Appl
43	26.6	26.3	7745	12	US-10-002-769-1	Sequence 1, Appl
44	26.4	26.1	2658	10	US-09-815-242-4035	Sequence 4035, Ap
45	26.2	25.9	1186	10	US-09-822-849A-398	Sequence 398, App

ALIGNMENTS

RESULT 1
US-09-861-289-36
Sequence 36, Application US/09861289
Patent No. US20020110897A1
GENERAL INFORMATION:
APPLICANT: Sherman, D.H.
APPLICANT: Liu, H.
APPLICANT: Xue, Y.
APPLICANT: Zhao, L.
TITLE OF INVENTION: DNA encoding methymycin and pikromycin
FILE REFERENCE: 600.438US1
CURRENT APPLICATION NUMBER: US/09/861,289
CURRENT FILING DATE: 2001-05-18
PRIOR APPLICATION NUMBER: 09/105,537
PRIOR FILING DATE: 1998-06-26
NUMBER OF SEQ ID NOS: 43
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 36
LENGTH: 4041
TYPE: DNA
ORGANISM: Streptomyces venezuelae
US-09-861-289-36

Query Match 38.2% Score 38.6; DB 10; Length 4041;
Best Local Similarity 61.4%; Pred. No. 0.0042;
Matches 62; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 1 GTGACGGCGGAGATCGCGCGCTGCTTCATGGGGCGCGAGCGCCGTGCGAT 60
DB 2851 GTGATGGCGGACGAGCGCGCTGCTCGGTCGACGTCGCCGAGAGAGTCCCGTCGAC 2910
QY 61 CGGCGCGTGTGCGACTTGCGCTGCACATCGCTCAGCGCGGT 101
DB 2911 CGCCCGCTGCGGAGATGCGCTTCGACCTCGACCGCGGT 2951

RESULT 2
US-09-861-289-5
Sequence 5, Application US/09861289
Patent No. US20020110897A1
GENERAL INFORMATION:
APPLICANT: Sherman, D.H.

```
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600 438US1
; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 36778
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-861-289-5
```

```
Query Match 38.2%; Score 38.6; DB 10; Length 36778;
Best Local Similarity 61.4%; Pred. No. 0.004;
Matches 62; Conservative 0; Mismatches 39; Indels 0; Gaps 0;
```

```
QY 1 GTGCAGGCGGAGATCGGCGCGTGTTCATATGGGCGCGGAGCGCCGCGCTCGAT 60
DB 34632 GTATGCGGCGAGCGCGCTCTCGTCTCGGCTCGACTCGCCCGAAGAGTCCCGCTGCAC 34691
QY 61 CGGCCGCTGTGCGACTTGGGCTCGACTCGCTCAGCGCGGT 101
DB 34692 CGGCCGCTGTGCGAGATCGGCTTCGACTCGCTCAGCGCGGT 34732
```

```
RESULT 3
US-09-861-289-32
; Sequence 32, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600 438US1
; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 32
; LENGTH: 11220
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-861-289-32
```

```
Query Match 33.5%; Score 33.8; DB 10; Length 11220;
Best Local Similarity 58.4%; Pred. No. 0.097;
Matches 59; Conservative 0; Mismatches 42; Indels 0; Gaps 0;
```

```
QY 1 GTGCAGGCGGAGATCGGCGCGTGTTCATATGGGCGCGGAGCGCCGCGCTCGAT 60
DB 10726 GTCCGTAACGACAGCTGGCGACCTCTGGGACAGGACCCCGGCGGTGGAGCTGGAG 10785
QY 61 CGGCCGCTGTGCGACTTGGGCTCGACTCGCTCAGCGCGGT 101
DB 10786 CGGCCCTTCCGCGACACCGGTTTGAAGCTGCTACCGCGCT 10826
```

```
RESULT 4
US-09-861-289-30
; Sequence 30, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
```

```
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600 438US1
; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 30
; LENGTH: 13842
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-861-289-30
```

```
Query Match 33.5%; Score 33.8; DB 10; Length 13842;
Best Local Similarity 58.4%; Pred. No. 0.096;
Matches 59; Conservative 0; Mismatches 42; Indels 0; Gaps 0;
```

```
QY 1 GTGCAGGCGGAGATCGGCGCGTGTTCATATGGGCGCGGAGCGCCGCGCTCGAT 60
DB 7501 GTACGGCGGAGCGCGCCGCGCTCTCGATGCTTCGCGGAGAGACGTCCGCGCGAC 7560
QY 61 CGGCCGCTGTGCGACTTGGGCTCGACTCGCTCAGCGCGGT 101
DB 7561 CGGCCCTTCAAGAGATCGGCTTCGACTCGCTCGCGCGGT 7601
```

```
RESULT 5
US-09-861-289-1
; Sequence 1, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600 438US1
; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 15872
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-861-289-1
```

```
Query Match 30.3%; Score 30.6; DB 10; Length 15872;
Best Local Similarity 56.4%; Pred. No. 0.79;
Matches 57; Conservative 0; Mismatches 44; Indels 0; Gaps 0;
```

```
QY 1 GTGCAGGCGGAGATCGGCGCGTGTTCATATGGGCGCGGAGCGCCGCGCTCGAT 60
DB 2822 GTCCGCGGACAGCTGCGCGCGCTCTCGATGAGACAGGACCCCGGCGGTGAGCGC 2881
QY 61 CGGCCGCTGTGCGACTTGGGCTCGACTCGCTCAGCGCGGT 101
DB 2882 CTCACCTTCAAGAGAGCTGGGCTTCGACTCCCTCATGTGCGGT 2922
```

```
RESULT 6
US-09-815-242-4056
; Sequence 4056, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
```

```
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4056
; LENGTH: 3189
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-815-242-4056
```

```
Query Match          30.1%; Score 30.4; DB 10; Length 3189;
Best Local Similarity 57.3%; Pred. No. 0.92; Mismatches 41; Indels 0; Gaps 0;
```

```
Matches 55; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 1 GTGAGGCGGAGATCGCGCGCTTTTCATGGGCGCGGAGCGCGCTGCGTCGAT 60
    ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1063 GTGTGTGTGTGATCCTTCTTCCTGCACAGCCTGGCGGCTCGATCCGCTGGCGGCG 1122
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 61 CGGCGGCTGTGCGACTTGGCGCTTCGACTCGCTCAG 96
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1123 GTGCCGCTGTGCGATGGCAGCACCCTTCGCGGTGATG 1158
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
RESULT 7
US-09-925-301-474
; Sequence 474, Application US/09925301
; Patent No. US20020052308A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA106
; CURRENT APPLICATION NUMBER: US/09/925,301
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05882
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1694
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 474
; LENGTH: 3209
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (427)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-301-474
```

```
Query Match          29.3%; Score 29.6; DB 10; Length 3209;
Best Local Similarity 56.0%; Pred. No. 1.6;
```

```
Matches 56; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

QY 2 TGCAGGCGGAGATCGCGCGGTGCTTCATGGGCGCGGAGCGCGCTGCGTCATC 61
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1215 TGTGTGAGGCCGTGGCACACTTCTGGGCATCCGCGACTTGCTGTCAACTGGACA 1274
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 62 GGCCGCTGTGCGACTTGGCGCTTCGATCGCTCAGCGCGGT 101
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1275 GCTCAGTGGCGGACCTGGCGCTGGACCTCGCATGACGCT 1314
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
RESULT 8
US-09-815-242-1227/c
; Sequence 1227, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haseldeck, Robert
; APPLICANT: Ohlson, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1227
; LENGTH: 99
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-815-242-1227
```

```
Query Match          29.1%; Score 29.4; DB 10; Length 99;
Best Local Similarity 56.8%; Pred. No. 1.9; Mismatches 41; Indels 0; Gaps 0;
```

```
Matches 54; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 2 TGCAGGCGGAGATCGCGCGGTGCTTCATGGGCGCGGAGCGCGCTGCGTCATC 61
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 99 TGTGTGTGTGATCCTTCTTCCTGCAGACTGGCGGCTGTGATATCCCGTGGCGCGG 40
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 62 GGCCGCTGTGCGACTTGGCGCTTCGATCGCTCAGCTCAGC 96
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 39 TGCCGCTGTGCGATCGGACACCTTCGCGGTGATG 5
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
RESULT 9
US-09-815-242-7980
; Sequence 7980, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haseldeck, Robert
; APPLICANT: Ohlson, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
```

```

? APPLICANT: Trawick, John D.
? APPLICANT: Carr, Grant J.
? APPLICANT: Yamamoto, Robert T.
? APPLICANT: Xu, H. Howard
? TITLE OF INVENTION: Identification of Essential Genes In
? TITLE OF INVENTION: Prokaryotes
? FILE REFERENCE: ELITRA.011A
? CURRENT APPLICATION NUMBER: US/09/815,242
? CURRENT FILING DATE: 2001-03-21
? PRIOR APPLICATION NUMBER: 60/191,078
? PRIOR FILING DATE: 2000-03-21
? PRIOR APPLICATION NUMBER: 60/206,848
? PRIOR FILING DATE: 2000-05-23
? PRIOR APPLICATION NUMBER: 60/207,727
? PRIOR FILING DATE: 2000-05-26
? PRIOR APPLICATION NUMBER: 60/242,578
? PRIOR FILING DATE: 2000-10-23
? PRIOR APPLICATION NUMBER: 60/253,625
? PRIOR FILING DATE: 2000-11-27
? PRIOR APPLICATION NUMBER: 60/257,931
? PRIOR FILING DATE: 2000-12-22
? PRIOR APPLICATION NUMBER: 60/269,308
? PRIOR FILING DATE: 2001-02-16
? NUMBER OF SEQ ID NOS: 14110
? SOFTWARE: FastSeq for Windows Version 4.0
? SEQ ID NO 7980
? LENGTH: 1413
? TYPE: DNA
? ORGANISM: Pseudomonas aeruginosa
? FEATURE:
? NAME/KEY: CDS
? LOCATION: (1)..(1413)
? OS-09-815-242-7980

```

```

Query Match      29.1%; Score 29.4; DB 10; Length 1413;
Best Local Similarity 58.6%; Pred. No. 1.8;
Matches 51; Conservative 0; Mismatches 36; Indels 0; Gaps 0

OY 15 CGCGCGCGCTGCTTCATCGGGCGCCGCGAGCGCGCTGCCGCTGATCGGCGCGCTCGGA 74
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 969 CCCCAGCGCGCTGTCGAGACGTACAGCGCGCGCGCGCTGCCGCGCTTCAGCGCTGATCAC 1028
OY 75 CTTGGGCGCTCGACATCGCTCAGCGCGCT 101
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1029 CCTGTGCTTCGCCCGCTGCTCTGTTGCT 1055

RESULT 10
US-09-919-172-93
: Sequence 93, Application US/09919172
: Patent No. US20020119463A1
: GENERAL INFORMATION:
: APPLICANT: Paris, Mary
: APPLICANT: Turner, Christopher M.
: TITLE OF INVENTION: PROSTATE CANCER MARKERS
: FILE REFERENCE: PA-0036 US
: CURRENT APPLICATION NUMBER: US/09/919,172
: CURRENT FILING DATE: 2001-07-30
: PRIOR APPLICATION NUMBER: 60/222,469
: PRIOR FILING DATE: 2000-07-28
: NUMBER OF SEQ ID NOS: 102
: SOFTWARE: PERL Program
: SEQ ID NO 93
: LENGTH: 1274
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: misc_feature
: OTHER INFORMATION: Incyte ID NO. US20020119463A1 002455.1
US-09-919-172-93

Query Match      28.7%; Score 29; DB 10; Length 1274;
Best Local Similarity 55.4%; Pred. No. 2.4;

```

	Matches	56;	Conservative	0;	Mismatches	45;	Indels	0;	Gaps	0;
Qy	1	GTGCGAGCCGAGATGGCGCGTGGTTTCATMGGGCGCGAGCGCCGTGCCGTGCAT	60							
Db	278	GGCGCGAGACGTGAGCTTGGGCGCACCTGGAGAGAGGGCGCGAGGCCGTGAGTCTTTTAG	337							
Qy	61	CGGCGCGCTGTGCGACTTTGGGCGCTGCATCTGCTACACGGCGGT	101							
Db	338	CTGCTGGGGCCGACTACGCGCGCCGGCACGGAGGGCGGACGT	378							

```

RESULT 11
US-09-861-289-15
; Sequence 15, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.4380S1
; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 15
; LENGTH: 1140
; type: DNA
; ORGANISM: Streptomyces venezuelae
; US-09-861-289-15

```

```

Query Match      27.7%; Score 28; DB 10; Length 1140;
Best Local Similarity 58.3%; Pred. No. 4.5;
Matches 49; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

QY      9  CGAGATCGCGCGCTGCTTCATGGGCGCGCGAGCGCGTGCCTGATCGGCGCT 68
          || | | | | | | | | | | | | | | | | | | | | | | | |
Db      282 CGCCAGCTGGCTCGCGGGGTCTCGCCACCGGCGCGACCCGCTGCCCGCGAGCGCACGA 341

QY      69  GTCGACTTGGGCGCTCGACTCGCT 92
          | | | | | | | | | | | | | | | | | |
Db      342 GGACCACCCACCGCTGAGCCGCT 365

RESULT 12
US-09-749-728B-12/c
; Sequence 12, Application US/09749728B
; Patent No. US20020142457A1
; GENERAL INFORMATION:
; APPLICANT: Umezawa, Akihiro
; APPLICANT: Hata, Jun-ichi
; APPLICANT: Fukuda, Keiichi
; APPLICANT: Ogawa, Satoshi
; APPLICANT: Sakurada, Kazuhiro
; APPLICANT: Goto, Satoshi
; APPLICANT: Yamada, Yoji
; TITLE OF INVENTION: THE CELL HAVING THE POTENTIALITY OF DIFFERENTIATION INTO CARDIO
; FILE REFERENCE: 00766.000043
; CURRENT APPLICATION NUMBER: US/09/749,728B
; CURRENT FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: H11-372826
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: PCT-JP00-01148
; PRIOR FILING DATE: 2000-02-28
; PRIOR APPLICATION NUMBER: PCT-JP00-07741
; PRIOR FILING DATE: 2000-11-02
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn Ver.2.0
; SEQ ID NO 12
; LENGTH: 1326

```

```
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; OTHER INFORMATION: (1)..(1329)
US-09-749-728B-12
```

```
Query Match
Best Local Similarity 55.0%; Pred. No. 4.5; Length 1326;
Matches 55; Conservative 0; Mismatches 45; Indels 0; Gaps 0;
```

```
QY 1 GTGAGCGCCGAGATCGCGCGCTTTTCATGAGGCGCCGCGAGCGCCGTCGAGT 60
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 470 GAGCCCGGAGAGCGCGCGCCGCTACTGCTGCGGCGCCGCGAGCGCCGCGCT 411
QY 61 CGGCGCTGTGCGACTTGGGCTTGCCTGACTCGCTACGCGG 100
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 410 CGGCGCGCTGTGTGTAGCGCGAGCTTCCGCGCGCGG 371
```

```
RESULT 13
US-09-861-289-3/C
; Sequence 3, Application US/09861289
```

```
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 13613
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-861-289-3
```

```
Query Match
Best Local Similarity 27.7%; Score 28; DB 10; Length 13613;
Matches 49; Conservative 0; Mismatches 35; Indels 0; Gaps 0;
```

```
QY 9 CGAGATCGCGCGCTGCTTCATGAGGCGCGCGAGCGCGCTGCGATCGCGCT 68
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 7688 CGCAGCTGTGCTGCGGTGTGCGCAGCGCGAGCCGCCCTGCCCTCGAGCGCACA 7629
QY 69 GTGCGACTTGGGCTCGACTGCT 92
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 7628 GGACCAACCCACCTTGAGCCGCT 7605
```

```
RESULT 14
US-09-967-477B-3
; Sequence 3, Application US/09967477B
; Patent No. US20020156254A1
; GENERAL INFORMATION:
; APPLICANT: Xiao Qiu
; APPLICANT: Haiping Hong
; TITLE OF INVENTION: FAD4, FAD5, FAD5-2, AND FAD6, NOVEL
; FILE REFERENCE: BNZ-001
; CURRENT APPLICATION NUMBER: US/09/967,477B
; CURRENT FILING DATE: 2002-04-16
; PRIOR APPLICATION NUMBER: 60/236,303
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: 60/297,562
; PRIOR FILING DATE: 2001-06-12
; NUMBER OF SEQ ID NOS: 8
```

```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1320
; TYPE: DNA
; ORGANISM: Thraustochytrium sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1320)
US-09-967-477B-3
```

```
Query Match
Best Local Similarity 27.3%; Score 27.6; DB 9; Length 1320;
Matches 51; Conservative 0; Mismatches 39; Indels 0; Gaps 0;
```

```
QY 12 GATCGCGCGCTGCTTCATGAGGCGCGCGAGCGCGCTGCGATGCGCGCTGTC 71
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 366 GATCCCGCACATGATTTTACCGCGTGTGAGATGTGCGCTTTCGCGCTGCTGTC 425
QY 72 GGACTTGGGCGCTCGACTGCTACGCGGCT 101
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 426 GCTCATGTCCAGGCTTCCGCGCGCTGCT 455
```

```
RESULT 15
US-09-960-352-8797
; Sequence 8797, Application US/09960352
```

```
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathalagan, Nagapan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 8797
; LENGTH: 446
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 38-LIB2809-003-Q1-E1-B6
US-09-960-352-8797
```

```
Query Match
Best Local Similarity 27.1%; Score 27.4; DB 10; Length 446;
Matches 52; Conservative 0; Mismatches 41; Indels 0; Gaps 0;
```

```
QY 2 TGCAGCGCGAGATCGCGCGCTGCTTCATGAGGCGCGCGAGCGCGCTGCGATC 61
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 76 TGCAGATGTTTCTTCACACGCTGTGCGCGAGCGCGCTGTGCGCGCGCTGTC 135
QY 62 GCGCGCTGTGCGACTTGGGCTCGACTGCTGCA 94
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 136 AGCTGACAGCTGATCTGTGCTGATCGCGCA 168
```

```
Search completed: November 6, 2002, 20:31:09
Job time : 46.125 secs
```


GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 6, 2002, 11:51:54 : Search time 33.5 Seconds
(without alignments)
924.608 Million cell updates/sec

Title: US-09-724-876-2_COPY_31170_31270
Perfect score: 101
Sequence: 1 gtgcagccgcgcagatcgcgcg.....ctcactcgtcaccgcggt 101

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapept 1.0

Searched: 441362 seqs, 1533381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA :
1: /cgn2_6/prodata/2/ina/5A.COMB.seq:*
2: /cgn2_6/prodata/2/ina/5B.COMB.seq:*
3: /cgn2_6/prodata/2/ina/6A.COMB.seq:*
4: /cgn2_6/prodata/2/ina/6B.COMB.seq:*
5: /cgn2_6/prodata/2/ina/PCTUS.COMB.seq:*
6: /cgn2_6/prodata/2/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	101	100.0	71989	4	US-09-443-501A-2 Sequence 2, Appl
2	99.4	98.4	68750	3	US-09-335-409-1 Sequence 1, Appl
3	99.4	98.4	68750	4	US-09-568-102-1 Sequence 1, Appl
4	99.4	98.4	68750	4	US-09-567-969-1 Sequence 1, Appl
5	99.4	98.4	68750	4	US-09-568-480-1 Sequence 1, Appl
6	99.4	98.4	68750	4	US-09-568-486-1 Sequence 1, Appl
7	99.4	98.4	68750	4	US-09-568-472-1 Sequence 1, Appl
8	99.4	98.4	68750	4	US-09-567-899-1 Sequence 1, Appl
9	40.4	40.0	80161	3	US-09-036-987A-1 Sequence 1, Appl
10	40.4	40.0	80161	4	US-09-370-700-1 Sequence 1, Appl
11	40.2	39.8	561	3	US-09-154-083-13 Sequence 13, Appl
12	38.6	38.2	4041	4	US-09-105-537-36 Sequence 36, Appl
13	38.6	38.2	36778	4	US-09-105-537-5 Sequence 5, Appl
14	38.6	38.2	38506	3	US-09-320-878-19 Sequence 19, Appl
15	38.6	38.2	50937	4	US-09-428-517-1 Sequence 1, Appl
16	37	36.6	43280	2	US-08-804-227C-1 Sequence 1, Appl
17	37	36.6	44377	2	US-08-804-227C-7 Sequence 7, Appl
18	37	36.6	44377	2	US-08-804-198-1 Sequence 1, Appl
19	35.4	35.0	1434	4	US-09-434-288-3 Sequence 3, Appl
20	35.4	35.0	4403765	4	US-09-103-840A-2 Sequence 2, Appl
21	33.8	33.5	11220	4	US-09-105-537-32 Sequence 32, Appl
22	33.8	33.5	13842	4	US-09-105-537-30 Sequence 30, Appl
23	33.8	33.5	33529	4	US-09-144-085-3 Sequence 3, Appl
24	32.2	31.9	11219	1	US-07-642-734C-1 Sequence 1, Appl
25	32.2	31.9	11219	3	US-08-439-009A-1 Sequence 1, Appl
26	31.2	30.9	23673	4	US-09-773-816-1 Sequence 1, Appl
27	31.2	30.9	441529	4	US-09-103-840A-1 Sequence 1, Appl

28	30.6	30.3	1681	4	US-09-434-288-7 Sequence 7, Appl
29	30.6	30.3	15872	4	US-09-105-537-1 Sequence 1, Appl
30	29.6	29.3	8460	1	US-08-469-005A-9 Sequence 9, Appl
31	29.6	29.3	8519	4	US-09-261-907-1 Sequence 1, Appl
32	29.6	29.3	20235	1	US-07-642-734C-3 Sequence 3, Appl
33	29.6	29.3	20235	3	US-08-439-009A-3 Sequence 3, Appl
34	28	27.7	1140	4	US-09-105-537-15 Sequence 15, Appl
35	28	27.7	1476	4	US-09-434-288-12 Sequence 12, Appl
36	28	27.7	3292	3	US-09-320-878-22 Sequence 22, Appl
37	28	27.7	13613	4	US-09-105-537-3 Sequence 3, Appl
38	28	27.7	28958	1	US-08-258-261B-6 Sequence 6, Appl
39	28	27.7	28958	1	US-08-456-337-6 Sequence 6, Appl
40	28	27.7	28958	1	US-08-457-342-6 Sequence 6, Appl
41	28	27.7	28958	1	US-08-457-342-6 Sequence 6, Appl
42	28	27.7	28958	1	US-08-458-076A-6 Sequence 6, Appl
43	28	27.7	28958	1	US-08-764-233A-4 Sequence 4, Appl
44	28	27.7	28958	1	US-08-457-335A-6 Sequence 6, Appl
45	28	27.7	28958	1	US-08-729-214-6 Sequence 6, Appl

ALIGNMENTS

RESULT 1
US-09-443-501A-2
Sequence 2, Application US/09443501A
Patent No. 6303342
GENERAL INFORMATION:
APPLICANT: Kosan Biosciences, Inc.
APPLICANT: Julien, Bryan
APPLICANT: Katz, Leonard
APPLICANT: Khosla, Chaitan
APPLICANT: Tang, Li
APPLICANT: Ziermann, Rainer
TITLE OF INVENTION: Recombinant Methods and Materials for Producing
FILE REFERENCE: 30062-20031.00
CURRENT APPLICATION NUMBER: US/09/443,501A
CURRENT FILING DATE: 1999-11-19
PRIOR APPLICATION NUMBER: US 60/130,560
PRIOR FILING DATE: 1999-04-22
PRIOR APPLICATION NUMBER: US 60/122,620
PRIOR FILING DATE: 1999-03-03
PRIOR APPLICATION NUMBER: US 60/119,386
PRIOR FILING DATE: 1999-02-10
PRIOR APPLICATION NUMBER: US 60/109,401
PRIOR FILING DATE: 1998-11-20
NUMBER OF SEQ ID NOS: 22
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 71989
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Synthetic construct
US-09-443-501A-2
Query Match 100.0%; Score 101; DB 4; Length 71989;
Best Local Similarity 100.0%; Pred. No. 5,1e-19;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GTGCAGCCGCGAGATCGCGCGCTTTCATGAGGCGCCGCGAGCGCGCGCTGCAT 60
Db 31170 GTGCAGCCGCGAGATCGCGCGCTTTCATGAGGCGCCGCGAGCGCGCGCTGCAT 31229
QY 61 CGGCGCTGTGCGACTGTGGCGCTGCATCGCTCAGCGCGGT 101
Db 31230 CGGCGCTGTGCGACTGTGGCGCTGCATCGCTCAGCGCGGT 31270
RESULT 2
US-09-335-409-1
Sequence 1, Application US/09335409

```
; Patent No. 6121029
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/335,409
; PRIORITY FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-335-409-1
```

```
Query Match          98.4%; Score 99.4; DB 3; Length 68750;
Best Local Similarity 99.0%; Pred. No. 1.4e-18;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 GTGCAGCCGAGATCGCGCGCTGCTTTCATGAGGCGCGCGAGCGCGCTGCCGTCGAT 60
|||||
DB 36782 GTGCAGCCGAGATCGCGCGCTGCTTTCATGAGGCGCGCGAGCGCGCTGCCGTCGAT 36841
```

```
QY 61 CGGCCGCTGTGGACTTGGGCTCGACTCGCTCAGCGCGGT 101
|||||
DB 36842 CGGCCGCTGTGGACTTGGGCTCGACTCGCTCAGCGCGGT 36882
```

```
RESULT 3
US-09-568-102-1
; Sequence 1, Application US/09568102
; Patent No. 6346404
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,102
; PRIORITY FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIORITY FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-102-1
```

```
Query Match          98.4%; Score 99.4; DB 4; Length 68750;
Best Local Similarity 99.0%; Pred. No. 1.4e-18;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 GTGCAGCCGAGATCGCGCGCTGCTTTCATGAGGCGCGCGAGCGCGCTGCCGTCGAT 60
|||||
DB 36782 GTGCAGCCGAGATCGCGCGCTGCTTTCATGAGGCGCGCGAGCGCGCTGCCGTCGAT 36841
```

```
QY 61 CGGCCGCTGTGGACTTGGGCTCGACTCGCTCAGCGCGGT 101
|||||
DB 36842 CGGCCGCTGTGGACTTGGGCTCGACTCGCTCAGCGCGGT 36882
```

```
RESULT 4
US-09-567-969-1
```

```
; Sequence 1, Application US/09567969
; Patent No. 6355457
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/567,969
; PRIORITY FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIORITY FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-567-969-1
```

```
Query Match          98.4%; Score 99.4; DB 4; Length 68750;
Best Local Similarity 99.0%; Pred. No. 1.4e-18;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 GTGCAGCCGAGATCGCGCGCTGCTTTCATGAGGCGCGCGAGCGCGCTGCCGTCGAT 60
|||||
DB 36782 GTGCAGCCGAGATCGCGCGCTGCTTTCATGAGGCGCGCGAGCGCGCTGCCGTCGAT 36841
```

```
QY 61 CGGCCGCTGTGGACTTGGGCTCGACTCGCTCAGCGCGGT 101
|||||
DB 36842 CGGCCGCTGTGGACTTGGGCTCGACTCGCTCAGCGCGGT 36882
```

```
RESULT 5
US-09-568-480-1
; Sequence 1, Application US/09568480
; Patent No. 6355458
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,480
; PRIORITY FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIORITY FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-480-1
```

```
Query Match          98.4%; Score 99.4; DB 4; Length 68750;
Best Local Similarity 99.0%; Pred. No. 1.4e-18;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 GTGCAGCCGAGATCGCGCGCTGCTTTCATGAGGCGCGCGAGCGCGCTGCCGTCGAT 60
|||||
DB 36782 GTGCAGCCGAGATCGCGCGCTGCTTTCATGAGGCGCGCGAGCGCGCTGCCGTCGAT 36841
```

```
QY 61 CGGCCGCTGTGGACTTGGGCTCGACTCGCTCAGCGCGGT 101
|||||
DB 36842 CGGCCGCTGTGGACTTGGGCTCGACTCGCTCAGCGCGGT 36882
```

```
RESULT 6
US-09-568-486-1
; Sequence 1, Application US/09568486
; Patent No. 6353459
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,486
; PRIORITY FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-486-1

Query Match          98.4%; Score 99.4; DB 4; Length 68750;
Best Local Similarity 99.0%; Pred. No. 1.4e-18;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTGCAGGCCGAGATCGCGCGCTTTCATGTGGGCGCCGAGCGCCGTCGCAT 60
Db 36782 GTGCAGGCCGAGATCGCGCGCTTTCATGTGGGCGCCGAGCGCCGTCGCAT 36841
QY 61 CGGCCGCTGTGCGACTTGGGCGCTCGACTCGCTCACGGCGGT 101
Db 36842 CGGCCGCTGTGCGACTTGGGCGCTCGACTCGCTCACGGCGGT 36882

RESULT 7
US-09-568-472-1
; Sequence 1, Application US/09568472
; Patent No. 6358719
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,472
; PRIORITY FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-472-1

Query Match          98.4%; Score 99.4; DB 4; Length 68750;
Best Local Similarity 99.0%; Pred. No. 1.4e-18;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTGCAGGCCGAGATCGCGCGCTTTCATGTGGGCGCCGAGCGCCGTCGCAT 60
Db 36782 GTGCAGGCCGAGATCGCGCGCTTTCATGTGGGCGCCGAGCGCCGTCGCAT 36841
QY 61 CGGCCGCTGTGCGACTTGGGCGCTCGACTCGCTCACGGCGGT 101
Db 36842 CGGCCGCTGTGCGACTTGGGCGCTCGACTCGCTCACGGCGGT 101
```

```
Db 36842 CGGCCGCTGTGCGACTTGGGCGCTCGACTCGCTCACGGCGGT 36882

RESULT 8
US-09-567-899-1
; Sequence 1, Application US/09567899
; Patent No. 6383787
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/567,899
; PRIORITY FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-567-899-1

Query Match          98.4%; Score 99.4; DB 4; Length 68750;
Best Local Similarity 99.0%; Pred. No. 1.4e-18;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTGCAGGCCGAGATCGCGCGCTTTCATGTGGGCGCCGAGCGCCGTCGCAT 60
Db 36782 GTGCAGGCCGAGATCGCGCGCTTTCATGTGGGCGCCGAGCGCCGTCGCAT 36841
QY 61 CGGCCGCTGTGCGACTTGGGCGCTCGACTCGCTCACGGCGGT 101
Db 36842 CGGCCGCTGTGCGACTTGGGCGCTCGACTCGCTCACGGCGGT 36882

RESULT 9
US-09-036-987A-1
; Sequence 1, Application US/09036987A
; Patent No. 6143526
; GENERAL INFORMATION:
; APPLICANT: Baltz, Richard H.
; APPLICANT: Broughton, Mary C.
; APPLICANT: Crawford, Kathryn P.
; APPLICANT: Madduri, Krishnamurthy
; APPLICANT: Merlo, Donald J.
; APPLICANT: Treadway, Patli J.
; APPLICANT: Turner, Jan R.
; APPLICANT: Waldron, Clive
; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
; TITLE OF INVENTION: Production
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Dow Agrosciences LLC Patent Department
; STREET: 9330 Zionsville Road
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: USA
; ZIP: 46268
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/036,987A
; FILING DATE: 09-MAR-1998
```

```

; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Stuart, Donald R
; REGISTRATION NUMBER: 28, 479
; REFERENCE/DOCKET NUMBER: 50, 608
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (317)337-4816
; TELEFAX: (317)337-4847
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 80161 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-09-036-987A-1

Query Match          40.0%; Score 40.4; DB 3; Length 80161;
Best Local Similarity 63.3%; Pred. No. 0.0069;
Matches 62; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 1 GTGCAGCCGAGATCGCGCGCTTTCATGTGGGCGCGCGCGCGCGCGCGAT 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 75463 GTCCGCGCGGATATCGCGTGTACTGGGCGACGCGAGTGCAGCATGTGACATCGAG 75522
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 61 CGGCGCGTGTGCGACTTGGGCGCTCGACTCGTCACGCGC 98
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 75523 AAGCCTTGGCGGAGCTGGGTTTCGACTCGTCGACGCGC 75560
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 10
US-09-370-700-1
; Sequence 1, Application US/09370700
; Patent No. 6274350
; GENERAL INFORMATION:
; APPLICANT: Baltz, Richard H
; APPLICANT: Broughton, Mary C
; APPLICANT: Crawford, Kathryn P
; APPLICANT: Madhuri, Krishnamurthy
; APPLICANT: Treadway, Patli J
; APPLICANT: Turner, Jan R
; APPLICANT: Waldron, Clive
; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
; FILE REFERENCE: 50489 Div1
; CURRENT APPLICATION NUMBER: US/09/370,700
; CURRENT FILING DATE: 1999-08-09
; EARLIER APPLICATION NUMBER: US 09/36987
; EARLIER FILING DATE: 1998-03-09
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: Patent Ver. 2.0
; SEQ ID NO 1
; LENGTH: 80161
; TYPE: DNA
; ORGANISM: Saccharopolyspora spinosa
; US-09-370-700-1

Query Match          40.0%; Score 40.4; DB 4; Length 80161;
Best Local Similarity 63.3%; Pred. No. 0.0069;
Matches 62; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 1 GTGCAGCCGAGATCGCGCGCTTTCATGTGGGCGCGCGCGCGCGCGCGAT 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 75463 GTCCGCGCGGATATCGCGTGTACTGGGCGACGCGAGTGCAGCATGTGACATCGAG 75522
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 61 CGGCGCGTGTGCGACTTGGGCGCTCGACTCGTCACGCGC 98
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 75523 AAGCCTTGGCGGAGCTGGGTTTCGACTCGTCGACGCGC 75560
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 11
US-09-154-083-13/C
; Sequence 13, Application US/09154083
; Patent No. 6150513
```

```

; GENERAL INFORMATION:
; APPLICANT: Wu, Kai
; TITLE OF INVENTION: Polyketide Synthase Enzymes and Recombinant DNA
; TITLE OF INVENTION: Constructs Therefor
; FILE REFERENCE: 30062-20022.00 Polyketide Synthase Enz
; CURRENT APPLICATION NUMBER: US/09/154,083
; CURRENT FILING DATE: 1998-09-16
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: Patent Ver. 2.0
; SEQ ID NO 13
; LENGTH: 561
; TYPE: DNA
; ORGANISM: Streptomyces hygroscopicus
; US-09-154-083-13

Query Match          39.8%; Score 40.2; DB 3; Length 561;
Best Local Similarity 62.4%; Pred. No. 0.008;
Matches 63; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY 1 GTGCAGCCGAGATCGCGCGCTTTCATGTGGGCGCGCGCGCGCGCGCTGCGAT 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 371 GTGGGCGACAGAGCGCGCGATCGCGCTCGGCGACAGGCTCCGAGCGCGCGCTGAGCGCGGAC 312
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 61 CGGCGCGTGTGCGACTTGGGCGCTCGACTCGTCACGCGCGT 101
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 311 AAGCGCTTCGGGAGATCGCGCTTCGACTCGTCGACGCGCGCT 271
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 12
US-09-105-537-36
; Sequence 36, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D. H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 36
; LENGTH: 4041
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
; US-09-105-537-36

Query Match          38.2%; Score 38.6; DB 4; Length 4041;
Best Local Similarity 61.4%; Pred. No. 0.021;
Matches 62; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 1 GTGCAGCCGAGATCGCGCGCTTTCATGTGGGCGCGCGCGCGCGCGCGAT 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2851 GTGATGCGGCGAGCGCGCTCCGCTCCGCTCGACTCGCCGGAAGAGTCCCGTCGAC 2910
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 61 CGGCGCGTGTGCGACTTGGGCGCTCGACTCGTCACGCGCGT 101
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2911 CGCCCGCTGCGGAGATCGGCTTCGACTCGTCGACGCGCGT 2951
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 13
US-09-105-537-5
; Sequence 5, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D. H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
```

```
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 36778
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-105-537-5
```

```
Query Match      38.2%; Score 38.6; DB 4; Length 36778;
Best Local Similarity 61.4%; Pred. No. 0.021;
Matches 62; Conservative 0; Mismatches 39; Indels 0; Gaps 0;
```

```
QY 1 GTGAGGCCGAGATCGGCGGCTTTCATGAGGCGCCCGAGCGCCGCTGCATGAT 60
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 34632 GTGATGCGGAGCGGCGCTCCGCTCCGGTCGACGTCGCCGAAGAGTCCCGTCGAC 34691
    || ||||| ||| | ||| ||||| ||| | | | | | | | | | | | | | |
```

```
QY 61 CGGCGCGCTGTCGACTGGGCGCTCGACTCGCTACGCGCGT 101
    || ||||| ||| | ||| ||||| ||| | | | | | | | | | | | | | |
Db 34692 CGCCCGCTGGGAGATCGGCTTCGACTCGCTGACCGCGCT 34732
```

```
RESULT 14
US-09-320-878-19
; Sequence 19, Application US/09320878A
; Patent No. 6117659
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETHLACH, Melanie C.
; APPLICANT: BETHLACH, Mary C.
; APPLICANT: MCDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 300622002120
; CURRENT APPLICATION NUMBER: US/09/320,878A
; CURRENT FILING DATE: 1999-05-27
; EARLIER APPLICATION NUMBER: CIP OF 09/141,908
; EARLIER FILING DATE: 1998-08-28
; EARLIER APPLICATION NUMBER: CIP OF 09/073,538
; EARLIER FILING DATE: 1998-05-06
; EARLIER APPLICATION NUMBER: CIP OF 08/846,247
; EARLIER FILING DATE: 1997-04-30
; EARLIER APPLICATION NUMBER: 60/119,139
; EARLIER FILING DATE: 1999-02-08
; EARLIER APPLICATION NUMBER: 60/100,880
; EARLIER FILING DATE: 1998-09-22
; EARLIER APPLICATION NUMBER: 60/087,080
; EARLIER FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 19
; LENGTH: 38506
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-320-878-19
```

```
Query Match      38.2%; Score 38.6; DB 3; Length 38506;
Best Local Similarity 61.4%; Pred. No. 0.021;
Matches 62; Conservative 0; Mismatches 39; Indels 0; Gaps 0;
```

```
QY 1 GTGAGGCCGAGATCGGCGGCTTTCATGAGGCGCCCGAGCGCCGCTGCATGAT 60
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 32774 GTGATGCGGAGCGGCGCTCCGCTCCGGTCGACGTCGCCGAAGAGTCCCGTCGAC 32833
```

```
QY 61 CGGCGCGCTGTCGACTGGGCGCTCGACTCGCTACGCGCGT 101
    || ||||| ||| | ||| ||||| ||| | | | | | | | | | | | | | |
Db 32834 CGCCCGCTGGGAGATCGGCTTCGACTCGCTGACCGCGCT 32874
```

```
RESULT 15
US-09-428-517-1
; Sequence 1, Application US/09428517
```

```
; Patent No. 6251636
; GENERAL INFORMATION:
; APPLICANT: Bethlach, Mary C.
; APPLICANT: Shah, Sanjay Krishnakant
; APPLICANT: MCDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 30062-20029.00
; CURRENT APPLICATION NUMBER: US/09/428,517
; CURRENT FILING DATE: 1999-10-28
; EARLIER APPLICATION NUMBER: 60/120,254
; EARLIER FILING DATE: 1999-02-16
; EARLIER APPLICATION NUMBER: 60/106,100
; EARLIER FILING DATE: 1998-10-29
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 1
; LENGTH: 50937
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Recombinant DNA
US-09-428-517-1
```

```
Query Match      38.2%; Score 38.6; DB 4; Length 50937;
Best Local Similarity 61.4%; Pred. No. 0.021;
Matches 62; Conservative 0; Mismatches 39; Indels 0; Gaps 0;
```

```
QY 1 GTGAGGCCGAGATCGGCGGCTTTCATGAGGCGCCCGAGCGCCGCTGCATGAT 60
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 39231 GTCCGTACCGAGCGAGCGCGCTGTCTGGGCGACGCGTCGCCGACGAGTCCGCGCGAG 39290
    || ||||| ||| | ||| ||||| ||| | | | | | | | | | | | | | |
QY 61 CGGCGCGCTGTCGACTGGGCGCTCGACTCGCTACGCGCGT 101
    || ||||| ||| | ||| ||||| ||| | | | | | | | | | | | | | |
Db 39291 CGGCGGTCAAGAGCGTGGGCTTCGACTCCCTACCGCGCTGT 39331
```

```
Search completed: November 6, 2002, 16:02:52
Job time : 217.625 secs
```


[illegible]

```

FT      /tag= d
FT      /note= "encodes ketide synthase (KS-Y) of the loading
FT      domain"
FT      misc_RNA
FT      3621..4661
FT      /tag= e
FT      /note= "encodes acyl transferase (AT) of the loading
FT      domain"
FT      misc_RNA
FT      4917..5810
FT      /tag= f
FT      /note= "encodes enoyl reductase (ER) of the loading
FT      domain, potentially involved in formation of the
FT      thiazole moiety"
FT      misc_RNA
FT      5856..6155
FT      /tag= g
FT      /note= "encodes acyl carrier protein (ACP) of the loading
FT      domain"
FT      CDS
FT      6260..10493
FT      /tag= h
FT      /label= epob_gene
FT      /note= "encodes module 1, the NRPS module"
FT      misc_RNA
FT      2031..3548
FT      /tag= i
FT      /note= "encodes condensation domain C2 of the NRPS
FT      module"
FT      misc_RNA
FT      2031..3548
FT      /tag= j
FT      /note= "encodes condensation domain C2 of the NRPS
FT      module"
FT      misc_RNA
FT      6861..6887
FT      /tag= k
FT      /note= "encodes heterocyclization signature sequence"
FT      misc_RNA
FT      6861..6887
FT      /tag= l
FT      /note= "encodes condensation domain C4 of the NRPS
FT      module"
FT      misc_RNA
FT      7358..7366
FT      /tag= m
FT      /note= "encodes condensation domain C7 (partial) of the
FT      NRPS module"
FT      misc_RNA
FT      7898..7921
FT      /tag= n
FT      /note= "encodes adenylation domain A1 of the NRPS module"
FT      misc_RNA
FT      7898..7921
FT      /tag= o
FT      /note= "encodes adenylation domain A1 of the NRPS module"
FT      misc_RNA
FT      8261..8308
FT      /tag= p
FT      /note= "encodes adenylation domain A3 of the NRPS module"
FT      misc_RNA
FT      8411..8422
FT      /tag= q
FT      /note= "encodes adenylation domain A4 of the NRPS module"
FT      misc_RNA
FT      8861..8905
FT      /tag= r
FT      /note= "encodes adenylation domain A6 of the NRPS module"
FT      misc_RNA
FT      8966..8983
FT      /tag= s
FT      /note= "encodes adenylation domain A7 of the NRPS module"
FT      misc_RNA
FT      9090..9179
FT      /tag= t
FT      /note= "encodes adenylation domain A8 of the NRPS module"
FT      misc_RNA
FT      9183..9992
FT      /tag= u
FT      /note= "encodes oxidation region for forming thiazole"
FT      misc_RNA
FT      10121..10138
FT      /tag= v
FT      /note= "encodes adenylation domain A10 of the NRPS
FT      module"
FT      misc_RNA
FT      10261..10306
FT      /tag= w
FT      /note= "encodes thiolation domain (PCP) of the NRPS
FT      module"
FT      CDS
FT      10639..16137
FT      /tag= x
FT      /label= epoc_gene
FT      /note= "encodes module 2"
FT      misc_RNA
FT      10654..12033
FT      /tag= y
FT      /note= "encodes KS2, the KS domain of module 2"
FT      misc_RNA
FT      12250..13287
FT      /tag= z
FT      /note= "encodes AT2, the AT domain of module 2"
FT      misc_RNA
FT      13327..13899
FT      /tag= aa
FT      /note= "encodes dehydratase (DH) 2, the DH domain of
FT      module 2"
FT      misc_RNA
FT      14962..15756
FT      /tag= ab
FT      /note= "encodes ketoreductase (KR) 2, the KR domain of
FT      module 2"
FT      misc_RNA
FT      15763..16008
FT      /tag= ac
FT      /note= "encodes ACP2, the ACP domain of module 2"
FT      CDS
FT      16134..37907
FT      /tag= ad
FT      /label= epod_gene
FT      /note= "encodes modules 3-6"
FT      misc_RNA
FT      16425..17606
FT      /tag= ae
FT      /note= "encodes KS3"
FT      misc_RNA
FT      17817..18857
FT      /tag= af
FT      /note= "encodes AT3"
FT      misc_RNA
FT      19581..20396
FT      /tag= ag
FT      /note= "encodes KR3"
FT      misc_RNA
FT      20424..20642
FT      /tag= ah
FT      /note= "encodes ACP3"
FT      misc_RNA
FT      20706..22082
FT      /tag= ai
FT      /note= "encodes KS4"
FT      misc_RNA
FT      22296..23336
FT      /tag= aj
FT      /note= "encodes AT4"
FT      misc_RNA
FT      24069..24647
FT      /tag= ak
FT      /note= "encodes KR4"
FT      misc_RNA
FT      24867..25151
FT      /tag= al
FT      /note= "encodes ACP4"
FT      misc_RNA
FT      25203..26576
FT      /tag= am
FT      /note= "encodes KS5"
FT      misc_RNA
FT      26793..27883
FT      /tag= an
FT      /note= "encodes AT5"
FT      misc_RNA
FT      27966..28574
FT      /tag= ao
FT      /note= "encodes DH5"
FT      misc_RNA
FT      29433..30287
FT      /tag= ap
FT      /note= "encodes ER5"
FT      misc_RNA
FT      30321..30869
FT      /tag= aq
FT      /note= "encodes KR5"
FT      misc_RNA
FT      31077..31373
FT      /tag= ar
FT      /note= "encodes ACP5"
FT      misc_RNA
FT      31440..32807
FT      /tag= as
FT      /note= "encodes KS6"
FT      misc_RNA
FT      33018..34067
FT      /tag= at
FT      /note= "encodes AT6"
FT      misc_RNA
FT      34107..34676
FT      /tag= au

```



```

FT      /note= "encodes DH6"
FT      35760..36641
FT      /tag= av
FT      /note= "encodes ER6"
FT      36705..37256
FT      /tag= aw
FT      /note= "encodes KR6"
FT      37470..37769
FT      /tag= ax
FT      /note= "encodes ACP6"
FT      37912..49308
FT      /tag= ay
FT      /label= epoE_gene
FT      /note= "encodes modules 7 and 8"
FT      38014..39375
FT      /tag= az
FT      /note= "encodes KS7"
FT      39589..40626
FT      /tag= ba
FT      /note= "encodes AT7"
FT      41341..41922
FT      /tag= bd
FT      /note= "encodes KR7"
FT      42181..42423
FT      misc_RNA

Query Match      100.0%; Score 101; DB 21; Length 71989;
Best Local Similarity 100.0%; Pred. No. 5.5e-18;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GTGCAGGCCGAGATCGCGCGCTTTCATGAGGGCGCGGAGCGCGCTGCCGTCGAT 60
Db      31170 GTGCAGGCCGAGATCGCGCGCTTTCATGAGGGCGCGGAGCGCGCTGCCGTCGAT 31229
QY      61 CGGCCGCTGTGCGACTTGGGCTCGACTCGCTCAGGGCGGT 101
Db      31230 CGGCCGCTGTGCGACTTGGGCTCGACTCGCTCAGGGCGGT 31270

RESULT 2
AAZ55887
ID      AAZ55887 standard; DNA; 68750 BP.
XX      AAZ55887;
XX      10-APR-2000 (first entry)
DE      Sorangium cellulosum 68.75 kb contig.
XX      Epothilone biosynthesis; type I polyketide synthase; taxol substitute;
KM      anticancer; ds.
XX      Sorangium cellulosum.
OS      Sorangium cellulosum.
FH      Key      Location/Qualifiers
FT      CDS      1..1826
FT      /tag= a
FT      /partial
FT      /product= "Partial Orf 1 protein (AAV58580)"
FT      /note= "No initiation codon given in the specification"
FT      complement (1900..3171)
FT      /tag= b
FT      /product= "Orf 2 protein (AAV58581)"
FT      3415..5556
FT      /tag= c
FT      /product= "Orf 3 protein (AAV58582)"
FT      complement (5612..5992)
FT      /tag= d
FT      /product= "Orf 4 protein (AAV58583)"
FT      6226..6675
FT      /tag= e
FT      /product= "Orf 5 protein (AAV58584)"
FT      7610..11875
FT      /tag= f

```

```

FT      /product= "Type I polyketide synthase, EPOS A
FT      (AAV58573)"
FT      11872..116104
FT      /tag= g
FT      /product= "Non-ribosomal peptide synthetase, EPOS P
FT      (AAV58574)"
FT      16251..21749
FT      /tag= h
FT      /product= "Type I polyketide synthase, EPOS B
FT      (AAV58575)"
FT      21746..43519
FT      /tag= i
FT      /product= "Type I polyketide synthase, EPOS C
FT      (AAV58576)"
FT      43524..54920
FT      /tag= j
FT      /product= "Type I polyketide synthase, EPOS D
FT      (AAV58577)"
FT      54935..62254
FT      /tag= k
FT      /product= "Type I polyketide synthase, EPOS E
FT      (AAV58578)"
FT      62369..63628
FT      /tag= l
FT      /product= "Cytochrome P450 oxygenase homologue, EPOS F
FT      (AAV58579)"
FT      63779..64333
FT      /tag= m
FT      /product= "Orf 6 protein (AAV58585)"
FT      complement (63853..64290)
FT      /tag= n
FT      /product= "Orf 7 protein (AAV58586)"
FT      64363..64920
FT      /tag= o
FT      /product= "Orf 8 protein (AAV58587)"
FT      complement (64287..64727)
FT      /tag= p
FT      /product= "Orf 9 protein (AAV58588)"
FT      65063..65767
FT      /tag= q
FT      /product= "Orf 10 protein"
FT      complement (65008..65874)
FT      /tag= r
FT      /product= "Orf 11 protein (AAV58590)"
FT      complement (65871..66338)
FT      /tag= s
FT      /product= "Orf 12 protein (AAV58591)"
FT      66667..67137
FT      /tag= t
FT      /product= "Orf 13 protein (AAV58592)"
FT      67334..68251
FT      /tag= u
FT      /product= "Orf 14 protein (AAV58593)"
FT      68346..68750
FT      /partial
FT      /tag= v
FT      /product= "Partial Orf 15 protein (AAV58594)"
FT      /note= "No termination codon given in the specification"
FT      WO9966028-A2.
XX      PD      23-DEC-1999.
XX      PF      16-JUN-1999; 99WO-EP04171.
XX      PR      18-JUN-1998; 98US-0099504.
XX      PR      24-SEP-1998; 98US-0101631.
XX      PR      05-FEB-1999; 99US-0118906.
XX      (NOVS ) NOVARTIS AG.
XX      PA      (NOVS ) NOVARTIS-ERFINDUNGEN VERR GRS MBH.
XX      PI      Schupp T, Ligon JM, Molnar I, Zirkle R, Goerlach J, Cyr D;

```

```

XX MPI: 2000-09741/08.
DR P-PSDB: AAY58573, AAY58574, AAY58575, AAY58576, AAY58577, AAY58578,
DR AAY58579, AAY58580, AAY58581, AAY58582, AAY58583, AAY58584,
DR AAY58585, AAY58586, AAY58587, AAY58588, AAY58590, AAY58591,
DR AAY58592, AAY58593, AAY58594.
XX
XX New isolated epothilone synthase genes, used for the recombinant
PT production of epothilone for use in cancer therapy.
XX
PS Claim 14: Page 87-104; 174pp: English.
XX
XX This sequence represents a 68.75 kb contig from Sorangium cellulosum
CC comprising 22 open reading frames (ORFs) and includes genes encoding
CC proteins involved in the biosynthesis of epothilones. Epothilones A and
CC B are 16-membered macrocyclic polyketides with an acylcysteine-derived
CC starter unit; polyketides being synthesised from two-carbon building
CC blocks, the beta-carbon of which always carries a keto group. Each round
CC of two-carbon addition is carried out by a complex of enzymes known as
CC the polyketide synthase in a manner similar to fatty acid biosynthesis.
CC EPOS A (AAY58573) and EPOS P (AAY58574) are involved in formation of
CC the thiazole ring formation of epothilones, and EPOS B, EPOS C, EPOS D
CC and EPOS E (AAY58575-58578) are involved in polyketide backbone
CC formation. EPO F (AAY58579) is an epothilone macrolactone oxidase, and
CC the protein Orf 3 (AAY58582) and Orf14 (AAY58593) are thought to be
CC involved in transport. Epothilones mimic the biological activity of
CC taxol, and may be substituted for taxol in cancer chemotherapeutic
CC compositions. Epothilones exhibit a much lower drop in potency against a
CC multiply drug-resistant cell line compared with taxol, and are
CC considerably less efficiently exported from such cells by the multidrug
CC resistance protein (MDR, or P-glycoprotein). Despite the potential of
CC epothilones as anticancer agents, they are problematical to produce on a
CC large scale. Epothilones are too complex for industrial scale chemical
CC synthesis, and Sorangium cellulosum is difficult to ferment, producing
CC poor yields of epothilones. The nucleic acids of the invention may be
CC used for the recombinant production of epothilones in a heterologous host
CC that is more amenable to fermentation.
XX
SQ Sequence 68750 BP; 9596 A; 22458 C; 25537 G; 11159 T; 0 other:

```

```

Query Match          98.4%; Score 99.4; DB 21; Length 68750;
Best Local Similarity 99.0%; Pred. No. 1.4e-17;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

OY 1 GTGCAGCCGAGATCGCGCGCTTTCATGAGGCGCGCGAGCCGCGCTGCAT 60
    |||||||
DB 36782 GTGCAGCGCGAGATCGCGCGCTTTCATGAGGCGCGCGAGCCGCGCTGCAT 36841
    |||||||
OY 61 CGGCGCGCTGTGCGACTTGGGCGCTCGACTCGCTCAGCGCGGT 101
    |||||||
DB 36842 CGGCGCGCTGTGCGACTTGGGCGCTCGACTCGCTCAGCGCGGT 36882
    |||||||

```

```

RESULT 3
AAV21187
ID AAV21187 standard; DNA; 53789 BP.
XX
XX AAV21187;
XX
XX 24-JUL-1998 (first entry)
XX
XX Amycolatopsis mediterranei rifamycin synthesis gene cluster fragment.
DE
XX
XX Amycolatopsis mediterranei rifamycin synthesis gene cluster;
KM polyketide synthase; actinomycete; ansamycin; ds.
XX
XX Amycolatopsis mediterranei.
OS
XX
XX Key Location/Qualifiers
FH 1825..15543
FT CDS
FT
FT /tag= a
FT /label= ORF_A
FT /product= "polyketide synthase"

```

```

FT CDS 15550..30759
FT /tag= b
FT /label= ORF_B
FT /product= "polyketide synthase"
FT CDS 30895..36060
FT /tag= c
FT /label= ORF_C
FT /product= "polyketide synthase"
FT CDS 36259..41325
FT /tag= d
FT /label= ORF_D
FT /product= "polyketide synthase"
FT CDS 41373..51614
FT /tag= e
FT /label= ORF_E
FT /product= "polyketide synthase"
FT CDS 51713..5293
FT /tag= f
FT /label= ORF_F
FT /product= "polyketide synthase"
XX
XX WO9807868-A1.
XX
XX 26-FEB-1998.
XX
XX 18-AUG-1997; 97WO-EP04495.
XX
XX 20-AUG-1996; 96EP-0810551.
XX
XX (NOVS ) NOVARTIS AG.
XX
XX Engel N, Schupp T, Toupet C;
XX
XX MPI: 1998-169172/15.
XX
XX P-PSDB: AAY52845-W52850.
XX
XX Amycolatopsis mediterranei rifamycin synthesis gene cluster - used
PT to produce rifamycin and rifamycin analogues
XX
XX
PS Claim 4; Page 53-102; 205pp: English.
XX
XX The present sequence represents a Amycolatopsis mediterranei rifamycin
CC synthesis gene cluster DNA fragment from the present invention. The
CC DNA fragment comprises a DNA region involved directly or indirectly
CC in the gene cluster responsible for rifamycin synthesis, including
CC the adjacent DNA regions to the right and left which, by reason of
CC their function in connection with rifamycin biosynthesis, qualify
CC as constituents of this rifamycin gene cluster, and functional
CC fragments, derivatives or constituents of these. The Amycolatopsis
CC mediterranei rifamycin synthesis gene cluster DNA fragment can be used
CC for producing rifamycin, rifamycin analogues or precursors. It can also
CC be used for inactivating or modifying genes involved in ansamycin or
CC rifamycin biosynthesis. The DNA can be used for constructing mutant
CC actinomycetes strains from which the natural rifamycin or ansamycin
CC biosynthesis gene cluster has been partly or completely deleted. The
CC DNA fragment can be used for assembling a library of polyketide
CC synthases, which can be used for assembling a library of polyketides.
CC A hybridisation probe of the invention can be used for identifying DNA
CC fragments involved in the biosynthesis of ansamycins.
XX
XX
SQ Sequence 53789 BP; 6707 A; 19183 C; 20504 G; 7395 T; 0 other:

```

```

Query Match          47.7%; Score 48.2; DB 19; Length 53789;
Best Local Similarity 67.3%; Pred. No. 0.00042;
Matches 68; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

```

```

OY 1 GTGCAGCCGAGATCGCGCGCTTTCATGAGGCGCGCGAGCCGCGCTGCAT 60
    || ||| || |||| ||||| || || ||||| ||||| || |||||
DB 25279 GTCCGGGCGCAGATCGCGCGCTTTCATGAGGCGCGAGCGAGCGCGCTGCAT 25338
    || ||| || |||| ||||| || || ||||| ||||| || |||||
OY 61 CGGCGCGCTGTGCGACTTGGGCGCTCGACTCGCTCAGCGCGGT 101
    || ||| || |||| ||||| || || ||||| ||||| || |||||
DB 25339 ACGGCGCTTCAAGAGCGCGCGCTTTCATGAGGCGCGCGCTGCAT 25379
    || ||| || |||| ||||| || || ||||| ||||| || |||||

```

```
RESULT 4
AAF88339
ID AAF88339 standard; DNA: 16767 BP.
XX
AC AAF88339;
XX
DT 28-AUG-2001 (first entry)
XX
DE S. spinosa DNA fragment encoding ORF22, SEQ ID 49.
XX
KW Forosamine; trimethylrharnose; polyketide synthase; biosynthesis;
KW spinosyn; polyketide aglycone; transgenic plant; insect resistance;
KW macrolide; insecticidal; polyketide synthase; ds.
XX
OS Saccharopolyspora spinosa.
XX
PN DE19957268-A1.
XX
PD 08-MAR-2001.
XX
PF 29-NOV-1999; 99DE-1057268.
XX
PR 27-AUG-1999; 99DE-1040596.
XX
PA (FARB ) BAYER AG.
XX
PI Eberz G, Moehrle V, Froede R, Velten R, Salas JA;
XX
DR WPI; 2001-267102/28.
XX
PT P-PSDB; AAB70969.
XX
PT New nucleic acid encoding enzymes for spinosyn biosynthesis, useful for
PT recombinant production of insecticidal spinosyns and their derivatives
XX
PS Claim 7a; Page 284-313; 354pp; German.
XX
CC This invention describes a novel method nucleic acid (I) and its encoded
CC polypeptide (II) containing at least one region that encodes an enzymatic
CC activity involved in biosynthesis of spinosyns. (I) are used (I) to
CC identify, inactivate or modulate genes involved in the biosynthesis of
CC (II); (II) to generate a library of polyketide synthases; (III) for
CC adding forosamine or trimethylrharnose to a spinosyn or polyketide
CC aglycone; and (IV) for recombinant production of the corresponding
CC enzymes, which are used for production of (II), their precursors or
CC derivatives, including production of transgenic plants that express (II)
CC and thus have increased resistance to insects. (I) are also useful as
CC markers for sequencing of the Saccharopolyspora spinosa genome. (II) are
CC macrolides with insecticidal, but not antibacterial, activity, and can
CC also be used to raise specific antibodies, useful for identifying
CC expression clones in a gene bank. Cells transformed with (I) may produce
CC (II) at significantly increased levels or produce new derivatives of
CC (II). This sequence, ORF 22, encodes an S. spinosa polyketide synthase.
XX
SQ Sequence 16767 BP; 2256 A; 4805 C; 6457 G; 3249 T; 0 other;

Query Match 40.0%; Score 40.4; DB 22; Length 16767;
Best Local Similarity 63.3%; Pred. No. 0.049;
Matches 62; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 1 GTGAGGCCGAGATCGCGCGCTCTTCATGGGGCCCGGAGCCCGTCGCGAT 60
   ||| ||||| ||||| ||| ||| ||||| ||| |||||
DB 15661 GTCCGCCCGCATATCGCGGTACTGGGGCAGCGAGTGTGACATCGAG 15720
   ||| ||| ||| ||| ||||| ||||| ||||| |||||
QY 61 CGGCGCGTGTGCGACTTGGGCGCTGACTCGCTCAGCGC 98
   ||| ||| ||| ||| ||||| ||||| ||||| |||||
DB 15721 AAGCTTTGGCGAGCTGGGTTTGCACCTCGCTGACGCG 15758

RESULT 5
AAF88314
```

```
ID AAF88314 standard; DNA: 25360 BP.
XX
AC AAF88314;
XX
DT 28-AUG-2001 (first entry)
XX
DE S. spinosa DNA fragment SEQ ID 3.
XX
KW Forosamine; trimethylrharnose; polyketide synthase; biosynthesis;
KW spinosyn; polyketide aglycone; transgenic plant; insect resistance;
KW macrolide; insecticidal; ds.
XX
OS Saccharopolyspora spinosa.
XX
PN DE19957268-A1.
XX
PD 08-MAR-2001.
XX
PF 29-NOV-1999; 99DE-1057268.
XX
PR 27-AUG-1999; 99DE-1040596.
XX
PA (FARB ) BAYER AG.
XX
PI Eberz G, Moehrle V, Froede R, Velten R, Salas JA;
XX
DR WPI; 2001-267102/28.
XX
PT New nucleic acid encoding enzymes for spinosyn biosynthesis, useful for
PT recombinant production of insecticidal spinosyns and their derivatives
XX
PS Claim 7; Page 49-58; 354pp; German.
XX
CC This invention describes a novel method nucleic acid (I) and its encoded
CC polypeptide (II) containing at least one region that encodes an enzymatic
CC activity involved in biosynthesis of spinosyns. (I) are used (I) to
CC identify, inactivate or modulate genes involved in the biosynthesis of
CC (II); (II) to generate a library of polyketide synthases; (III) for
CC adding forosamine or trimethylrharnose to a spinosyn or polyketide
CC aglycone; and (IV) for recombinant production of the corresponding
CC enzymes, which are used for production of (II), their precursors or
CC derivatives, including production of transgenic plants that express (II)
CC and thus have increased resistance to insects. (I) are also useful as
CC markers for sequencing of the Saccharopolyspora spinosa genome. (II) are
CC macrolides with insecticidal, but not antibacterial, activity, and can
CC also be used to raise specific antibodies, useful for identifying
CC expression clones in a gene bank. Cells transformed with (I) may produce
CC (II) at significantly increased levels or produce new derivatives of
CC (II). This sequence represents a genomic DNA fragment of the S. spinosa
CC genome which contains the coding regions for proteins involved in
CC forosamine, trimethylrharnose and polyketide synthase biosynthesis.
XX
SQ Sequence 25360 BP; 3832 A; 9143 C; 8354 G; 4031 T; 0 other;

Query Match 40.0%; Score 40.4; DB 22; Length 25360;
Best Local Similarity 63.3%; Pred. No. 0.048;
Matches 62; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 1 GTGAGGCCGAGATCGCGCGCTCTTCATGGGGCCCGGAGCCCGTCGCGAT 60
   ||| ||||| ||||| ||| ||| ||||| ||| |||||
DB 92 GTCCGCCCGCATATCGCGGTACTGGGGCAGCGAGTGTGACATCGAG 151
   ||| ||| ||| ||| ||||| ||||| ||||| |||||
QY 61 CGGCGCGTGTGCGACTTGGGCGCTGACTCGCTCAGCGC 98
   ||| ||| ||| ||| ||||| ||||| ||||| |||||
DB 152 AAGCTTTGGCGAGCTGGGTTTGCACCTCGCTGACGCG 189

RESULT 6
AAF88317
ID AAF88317 standard; DNA: 29736 BP.
XX
AC AAF88317;
```

XX 28-AUG-2001 (first entry)
 DT
 XX
 DE S. spinosa DNA fragment SEQ ID 6.
 XX
 XX
 KM Forsamine; trimethylrhinose; polyketide synthase; biosynthesis;
 KM spinosyn; polyketide aglycone; transgenic plant; insect resistance;
 KW macrolide; insecticidal; ds.
 XX
 XX Saccharopolyspora spinosa.
 OS
 PN DE19957268-A1.
 PD
 XX 08-MAR-2001.
 XX
 PF 29-NOV-1999; 99DE-1057268.
 XX
 PR 27-AUG-1999; 99DE-1040596.
 XX
 PA (FARB) BAYER AG.
 XX
 PI Eberz G, Moehrl V, Froede R, Velten R, Salas JA;
 XX
 DR WPI; 2001-267102/28.
 XX
 PT New nucleic acid encoding enzymes for spinosyn biosynthesis, useful for
 PT recombinant production of insecticidal spinosyns and their derivatives
 PT
 PT -
 XX
 PS Claim 7; Page 92-102; 354pp; German.
 XX
 CC This invention describes a novel method nucleic acid (I) and its encoded
 CC polypeptide (II) containing at least one region that encodes an enzymatic
 CC activity involved in biosynthesis of spinosyns. (I) are used (i) to
 CC identify, inactive or modulate genes involved in the biosynthesis of
 CC (II); (ii) to generate a library of polyketide synthases; (iii) for
 CC adding forosamine or trimethylrhinose to a spinosyn or polyketide
 CC aglycone; and (iv) for recombinant production of the corresponding
 CC enzymes, which are used for production of (II), their precursors or
 CC derivatives, including production of transgenic plants that express (II)
 CC and thus have increased resistance to insects. (I) are also useful as
 CC markers for sequencing of the Saccharopolyspora spinosa genome. (II) are
 CC macrolides with insecticidal, but not antibacterial, activity, and can
 CC also be used to raise specific antibodies, useful for identifying
 CC expression clones in a gene bank. Cells transformed with (I) may produce
 CC (II) at significantly increased levels or produce new derivatives of
 CC (II). This sequence represents a genomic DNA fragment of the S. spinosa
 CC genome which contains the coding regions for proteins involved in
 CC forosamine and trimethylrhinose biosynthesis.
 CC
 SQ Sequence 29736 BP; 4401 A; 10346 C; 10080 G; 4909 T; 0 other;
 Query Match 40.0%; Score 40.4; DB 22; Length 29736;
 Best Local Similarity 63.3%; Pred. No. 0.048; Mismatches 36; Indels 0; Gaps 0;
 Matches 62; Conservative 0; Mismatches 36; Indels 0; Gaps 0;
 QY 1 GTGCGACCGCGAGATCGCGCGTCTTCATGGGCGCGCGCCGCCGCTGCAT 60
 ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 4468 GTGCGGCGCGATATCGCGGCTACGTGGGCGACGCGAGTCGATGTGACATCGAG 4527
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 QY 61 CGGCCGCTGTGGACTTGGGCGCTCGACTCGCTACGCGC 98
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 DB 4528 AAGCCTTGTGGCGAGCTGGGTTTCGACTCGCTACGCGC 4565
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 RESULT 7
 AA21501
 ID AA21501 standard; DNA; 80161 BP.
 XX
 AC AA21501;
 XX
 DT 01-DEC-1999 (first entry)
 XX

DE DNA fragment of Saccharopolyspora spinosa containing biosynthetic genes.
 XX
 XX Spinosyn biosynthetic enzyme; open reading frame; ORF; insecticidal
 KM macrolides; arachnid; nematode; insect; polyketide; polyketide synthase;
 KM PKS; extender module; initiator module; acyl transferase domain; AT;
 KM acyl carrier protein; ACP; beta-ketosynthase domain; KS; KR;
 KM dehydratase domain; DH; enoyl reductase domain; ER; beta-ketoreductase;
 KW insecticide; ss.
 XX
 XX Saccharopolyspora spinosa.
 OS
 PH
 XX
 PH Key
 FT
 FT complement (1135..1971)
 FT
 FT /tag= a
 FT /product= ORFL16
 FT /note= "Protein involved in transcription control"
 FT 2024..2791
 FT
 FT /tag= b
 FT /product= ORFL15
 FT /note= "keto acyl reductase"
 FT complement (3416..4165)
 FT
 FT /tag= c
 FT /product= spns
 FT /note= "Spinosyn biosynthesis protein S"
 FT complement (4168..5325)
 FT
 FT /tag= d
 FT /product= spnr
 FT /note= "Spinosyn biosynthesis protein R"
 FT complement (5363..6751)
 FT
 FT /tag= e
 FT /product= spnq
 FT /note= "Spinosyn biosynthesis protein Q"
 FT 7083..8450
 FT
 FT /tag= f
 FT /product= spnp
 FT /note= "Spinosyn biosynthesis protein P"
 FT 8967..10427
 FT
 FT /tag= g
 FT /product= spno
 FT /note= "Spinosyn biosynthesis protein O"
 FT 10436..11434
 FT
 FT /tag= h
 FT /product= spnh
 FT /note= "Spinosyn biosynthesis protein N"
 FT complement (11530..12492)
 FT
 FT /tag= i
 FT /product= spnm
 FT /note= "Spinosyn biosynthesis protein M"
 FT complement (12696..13547)
 FT
 FT /tag= j
 FT /product= spnl
 FT /note= "Spinosyn biosynthesis protein L"
 FT complement (13592..14785)
 FT
 FT /tag= k
 FT /product= spnk
 FT /note= "Spinosyn biosynthesis protein K"
 FT complement (14799..16418)
 FT
 FT /tag= l
 FT /product= spnj
 FT /note= "Spinosyn biosynthesis protein J"
 FT 16536..17743
 FT
 FT /tag= m
 FT /product= spni
 FT /note= "Spinosyn biosynthesis protein I"
 FT complement (17749..18501)
 FT
 FT /tag= n
 FT /product= spnh
 FT /note= "Spinosyn biosynthesis protein H"
 FT complement (18541..19713)
 FT
 FT /tag= o
 FT /product= spng
 FT /note= "Spinosyn biosynthesis protein G"
 FT 20168..20995
 FT
 FT CDS

FT	/tag= p
FT	/product= spn ^f
FT	/note= "Spinosyn biosynthesis protein F"
FT	2111..28898
FT	/tag= q
FT	/product= spna
FT	/note= "Spinosyn biosynthesis protein A"
FT	/transl_except= (Pos:26940..26942, aa:Pro)
FT	21126..22379
FT	/tag= r
FT	/note= "Beta-ketosynthase domain (KS1) : part of the
FT	initiator module"
FT	22692..23669
FT	/tag= s
FT	/note= "Acyl transferase domain (AT1) : part of the
FT	initiator module"
FT	23793..24041
FT	/tag= t
FT	/note= "Acyl carrier protein domain (ACPI) : part of the
FT	initiator module"
FT	24102..25349
FT	/tag= u
FT	/note= "Beta-ketosynthase domain (KS1) : part of
FT	extender module 1"
FT	25683..26684
FT	/tag= v
FT	/note= "Acyl transferase domain (AT1) : part of
FT	extender module 1"
FT	27582..28121
FT	/tag= w
FT	/note= "Beta-ketoreductase domain (KR1) : part of
FT	extender module 1"
FT	28404..28649
FT	/tag= x
FT	/note= "Acyl carrier protein domain (ACPI) : part of
FT	extender module 1"
FT	38916..35374
FT	/tag= y
FT	/product= spnc
FT	/note= "Spinosyn biosynthesis protein C"
FT	29024..30295
FT	/tag= z
FT	/note= "Beta-ketosynthase domain (KS2) : part of
FT	extender module 2"
FT	30629..31621
FT	/tag= aa
FT	/note= "Acyl transferase domain (AT2) : part of
FT	extender module 2"
FT	31697..32254
FT	/tag= ab
FT	/note= "Dehydratase domain (DH2) : part of extender
FT	module 2"
FT	33035..34072
FT	/tag= ac
FT	/note= "Enoyl reductase domain (ER2) : part of
FT	extender module 2"
FT	34082..34621
FT	/tag= ad
FT	/note= "beta-ketoreductase domain (KR2) : part of
FT	extender module 2"
FT	34886..30295
FT	/tag= ae
FT	/note= "Acyl carrier protein domain (ACP2) : part of
FT	extender module 2"
FT	35419..44931
FT	/tag= af
FT	/product= spnc
FT	/note= "Spinosyn biosynthesis protein C"
FT	35518..36786
FT	/tag= ag
FT	/note= "beta-ketosynthase domain (KS3) : part of
FT	extender module 3"
FT	37108..38097

Query Match	Best Local Similarity	40.0%: Score 40.4; DB 20; Length 80161;
Matches 62; Conservative	0; Mismatches 36; Indels 0; Gaps 0;	
QY 1 GTGCAGGCCGAGATCGCGCGCTGCTTTTCATGAGGGCGCCGAGAGCGCCGTCGCCGTCGAT 60		
DB 75463 GTCCGCGCGCATATCGCGGTGACTGAGGCGACGCGCAGATGAGCGAGTGGACATCGAG 75522		
OY 61 CGGCGCGTGTGCGATCTGGCGCTGACATCGCTCAGCGC 98		
DB 75523 AAGCCTTTGGCCGAGCTGGTTTCAGTCGCTGACGCG 75560		
RESULT 8		
AAA89135/C		
ID AAA89135 standard; DNA; 561 BP.		
XX AAA89135;		
XX		
DT 19-MAR-2001 (first entry)		
XX		
DE Polyketide synthase derived sequence 34-119.1.T7 DNA.		
XX		
XX Polyketide synthase; PKS; cosmid 34-119; FK-520; antibiotic;		

KW		Immunosuppressive; neuroimmunophilin; ds.
XX		
OS	Streptomyces hygroscopicus.	
XX		
FH	Key	Location/Qualifiers
FT	CDS	complement (3..560)
FT		/tag= a
FT		/partial
FT		/note= "Includes in-frame stop codon at 516..518"
PN	US6150513-A.	
XX		
PD	21-NOV-2000.	
XX		
PF	16-SEP-1998;	98US-0154083.
XX		
PR	16-SEP-1998;	98US-0154083.
XX		
PA	(KOSA-) KOSAN BIOSCIENCES INC.	
XX		
P1	Wu K;	
DR	WPI: 2001-049104/06.	
DR	P-PSDB: AAB19902.	
XX		
P7	Novel nucleic acid encoding polyketide synthase useful for producing recombinant polyketide synthase for making novel polyketides for use as antibiotics, immunosuppressant or as neuroimmunophilin -	
XX		
PS	Example; Column 23-24; 30pp; English.	
XX		
CC	The present sequence is that of a polyketide synthase (PKS)-derived coding sequence designated 34-119.1.t7. The DNA was isolated from cosmid 34-183 of an EK-520-producing strain of Streptomyces hygroscopicus following restriction digestion. It does not comprise a full-length coding sequence. The 186-amino acid translated sequence is given in AAB19902. An isolated nucleic acid comprising a coding sequence for a PKS enzyme identical to, or isolated from, a PKS coding sequence contained within a cosmid selected from the cosmids 34-183, 34-122 and 34-126 is claimed. Such nucleic acids, when used alone or in combination with other PKS domain coding sequences, are useful in the construction of recombinant vectors that encode PKS enzymes which can be used to make novel polyketides having applications in medicine, agriculture and animal health. The polyketides may have e.g. antibiotic, immunosuppressive or neuroimmunophilin activities.	
CC		
CC		
CC		
S0	Sequence 561 BP; 83 A; 230 C; 189 G; 59 T; 0 other:	
Query Match	39.8%; Score 40.2; DB 22; Length 561;	
Best Local Similarity	62.4%; Pred. No. 0.061;	
Matches	63; Conservative 0; Mismatches 38; Indels 0; Gaps	
OY	1 GTGCAGGGCGGAGATGCGCGCGTGTTCATATGCGCGCGCAGAGCGCGCTGCCGTCGAT 60	
DB	371 GTTCGGCGACGAGCGCGCATTCGCTCGCGCACGCGCTCGAAGCGCGCTGGACGCGCAC 312	
OY	61 CGCGCGCTGTGCGACTTGAGCCCTGCATCTGCATCGATCAGCGCGGT 101	
DB	311 AAGCGCTTCCGGAGATCGGCTTCGACATCGCTGACGCGCGCT 271	
RESULT 9		
ID	AA287300	
AC	AA287300 standard; DNA; 4041 BP.	
XX		
AC	AA287300;	
XX		
DT	05-JUN-2000 (first entry)	
XX		
DE	S. venezuelae macrolide biosynthetic gene pikATV, SEQ ID NO:36.	
XX		
DE	Desosamine biosynthesis: macrolide; polyketide; methymycin; pikromycin;	

KW		neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
KM		bipolymer; antibiotic; chemotherapeutic; immunosuppressant; aschma;
KX		chronic obstructive pulmonary disease; respiratory inflammation;
XX		hypercholesterolaemia; crop protection agent; ds.
OS		Streptomyces venezuelae ATCC15439.
PH	Key	Location/Qualifiers
FT	CDS	1..4041
ET		/tag= a
PT		/product= "PlkAIV"
PN		MO20000620-A2.
XX		
PD		06-JAN-2000.
XX		
PF		25-JUN-1999; 99WO-US14398.
PR		26-JUN-1998; 98US-O105537.
XX		
PA		(MINU) UNIV MINNESOTA.
PI		Sherman DH, Liu H, Xue Y, Zhao L;
DR		WPI: 2000-160679/14.
XX		P-PDB: AAT771193.
PT		Desosamine and macrolide biosynthetic gene clusters, useful for, e.g.
PS		synthesis of methymycin and pikromycin -
XX		
Claim 15;	Page 422-424;	438bp; English.
XX		
CC	The invention relates to an isolated and purified nucleic acid segment	
CC	comprising a desosamine biosynthetic gene cluster, a fragment or its	
CC	biologically active variant, where the nucleic acid sequence is not	
CC	derived from the erc gene cluster of Saccharopolyspora erythraea or	
CC	Streptomycis antibiotics. The invention also relates to a macrolide	
CC	biosynthetic gene cluster, or fragments thereof. The macrolide	
CC	biosynthetic gene cluster encodes proteins which synthesise methymycin,	
CC	pikromycin, neomethymycin, narbomycin or a combination of these	
CC	compounds. Recombinant or augmented cells comprising the desosamine	
CC	and/or macrolide biosynthetic gene clusters are useful for the	
CC	production of biologically active macrolides. The macrolide biosynthetic	
CC	proteins are useful for synthesis of methymycin, pikromycin,	
CC	neomethymycin and narbomycin. The alternative termination of polyketide	
CC	synthesis may be useful to prepare novel antibiotics and	
CC	polyhydroxyalkanoate (PHA) monomers. The compounds produced by the	
CC	recombinant host cells are useful as biopolymers, e.g., in packaging or	
CC	biomedical applications, to engineer PHA monomer synthases or to prepare	
CC	immunologically active agents, such as chemotherapeutics,	
CC	immunosuppressants, agents to treat asthma, chronic obstructive pulmonary	
CC	disease as well as other diseases involving respiratory inflammation,	
CC	cholesterol-lowering agents or macrolide-based antibiotics which are	
CC	active against a variety of organisms, e.g., bacteria, including	
CC	multi-drug resistant pneumococci and other respiratory pathogens, as well	
CC	as viral parasitic pathogens, or as crop protection agents (e.g.,	
CC	fungicides or insecticides) via expression of polyketides in plants.	
CC	Sequences AA87295-887302 represent macrolide biosynthetic genes from	
CC	Streptomyces venezuelae ATCC 15439, which encode proteins	
CC	AAT771190-177197.	
XX		
SO	Sequence 4041 BP; 547 A; 1572 C; 1393 G; 529 T; 0 other:	
	Query Match	38.2%; Score 38.6; DB 21; Length 4041;
	Best Local Similarity	61.4%; Pred. No. 0.15;
	Matches	62; Conservative 0; Mismatches 39; Indels 0; Gaps 0;
OY	1 GTGACAGCGCAGATCGCGCGCTGTTCATGGAGCGCGAGCGCGCTGCCTGCAT	60
Db	2851 GTGATGCGGAGAGCGGCTCGTGTCTCGGTCGACTGCCCGGAAGAGATTCCCCTGCAC	2910
OY	61 CGGCGCTGTGCGACTTGCGGCTCTGCATCTGCTCACGCGCGT	101

DB 2911 CGCCCGTCGCGAGATCGGCTTCGACTCGTCGACCGCGT 2951
RESULT 10
ID AAD17185 standard; DNA; 27541 BP.
XX
AC AAD17185;
XX
DT 29-NOV-2001 (first entry)
XX
DE Streptomyces noursei nys2 DNA of nystatin PKS gene cluster.
XX
KM Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
XX antifungal; antibiotic; nys2; ds.
OS Streptomyces noursei.
XX
FH Key Location/Qualifiers
FT CDS complement (454..1191)
FT /tag= a
FT /product= "NysF protein"
FT /note= "CDS does not include start codon"
FT /tag= b
FT /product= "NysG protein"
FT complement (3070..4824)
FT /tag= c
FT /note= "CDS does not include start codon"
FT /product= "NysH protein"
FT 5122..6156
FT /tag= d
FT /product= "NysD3 protein"
FT 6338..27541
FT /tag= e
FT /product= "NysI partial protein"
FT /note= "CDS does not include stop codon"
XX
PN WO200159126-A2.
XX
PD 16-AUG-2001.
XX
PF 08-FEB-2001; 2001WO-GB00509.
XX
PR 08-FEB-2000; 2000GB-0002840.
PR 10-APR-2000; 2000GB-0008786.
PR 14-APR-2000; 2000GB-0009387.
XX
XX (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
PA (SNPE) SINTEF STEFTELSEN IMD TEK FORSK.
PA (ALPH-) ALPHARMA AS.
PA (SINV-) SINVENT AS.
PA (DZIE/) DZIEGLEWSKA H.
PA (ZOTC/) ZOTCHEV S B.
PA (SEKU/) SEKUROVA O N.
PA (FJAE/) FJAEVRIK E.
PA (BRAU/) BRAUTASET T.
PA (STRO/) STROM A R.
XX
XX Zotchev SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AF;
PI Valla S, Ellingsen TE, Sletta H, Gulliksen O;
XX
DR MPI: 2001-557614/62.
DR P-PSDB: AAE10138, AAE10139, AAE10140, AAE10141, AAE10142.
XX
XX New nystatin polyketide synthase polynucleotides and polypeptides,
PT useful as antibiotics and antifungals -
XX
PS Claim 2; Page 151-166; 266pp; English.
XX
XX The present invention relates to the cloning and sequencing of the gene
CC cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC involved in the biosynthesis of the macrolide antibiotic nystatin.

CC The nystatin PKS is useful as antifungal antibiotics. The present
CC sequence is a Streptomyces noursei nys2 DNA of nystatin PKS gene cluster.
XX
SQ Sequence 27541 BP; 3517 A; 10766 C; 9529 G; 3729 T; 0 other;
XX
Query Match 38.2%; Score 38.6; DB 22; Length 27541;
Best local Similarity 61.4%; Pred. No. 0.14;
Matches 62; Conservative 0; Mismatches 39; Indels 0; Gaps 0;
OY 1 GTCCAGCCGAGATCGCGCGCTTTCATGAGCGCGCGAGCGCGCTCGAT 60
DB 16145 GTCCGACCGAGTCCGCCACCGCTTCGCGCCACAGAGCTCCCGACCGCGAGGCG 16204
OY 61 CGCGCGCTTCGACTTGGGCGCTTCGACTCGCTCAGCGCGGT 101
DB 16205 CGCGCTTCGCGAGCTCGGCTTCGACTCGCTCAGCGCGGT 16245
RESULT 11
ID AA287318 standard; DNA; 36778 BP.
XX
AC AA287318;
XX
DT 05-JUN-2000 (first entry)
XX
DE S. venezuelae pik (macrolide biosynthesis) gene cluster.
XX
XX Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;
KM neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
KM biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;
KM chronic obstructive pulmonary disease; respiratory inflammation;
KM hypercholesterolaemia; crop protection agent; ds.
XX
OS Streptomyces venezuelae ATCC15439.
XX
FH Key Location/Qualifiers
FT CDS 1742..15583
FT /tag= a
FT /product= "pik gene cluster protein #1 (AAV77200)"
FT 15688..26907
FT /tag= b
FT /product= "pik gene cluster protein #2 (AAV77201)"
FT 26991..31679
FT /tag= c
FT /product= "pik gene cluster protein #3 (AAV77202)"
FT 31782..35822
FT /tag= d
FT /product= "pik gene cluster protein #4 (AAV77203)"
FT 35819..36664
FT /tag= e
FT /product= "pik gene cluster protein #4 (AAV80997)"
XX
XX WO200000620-A2.
XX
PD 06-JAN-2000.
XX
PF 25-JUN-1999; 99WO-US14398.
XX
PR 26-JUN-1998; 98US-0105537.
XX
XX (MING) UNIV MINNESOTA.
PI Sherman DH, Liu H, Xue Y, Zhao L;
XX
DR MPI: 2000-160679/14.
DR P-PSDB: AAV77200, AAV77201, AAV77202, AAV77203, AAV80997.
XX
XX Desosamine and macrolide biosynthetic gene clusters, useful for, e .g.
PT synthesis of methymycin and pikromycin -
XX
PS Disclosure; Figure 31; 438pp; English.
XX

CC The invention relates to an isolated and purified nucleic acid segment
CC comprising a desosamine biosynthetic gene cluster, a fragment or its
CC biologically active variant, where the nucleic acid sequence is not
CC derived from the eryC gene cluster of *Saccharopolyspora erythraea* or
CC streptomycetes antibiotics. The invention also relates to a macrolide
CC biosynthetic gene cluster, or fragments thereof. The macrolide
CC biosynthetic gene cluster encodes proteins which synthesise methymycin,
CC pikromycin, neomethymycin, narbomycin or a combination of these
CC compounds. Recombinant or augmented cells comprising the desosamine
CC and/or macrolide biosynthetic gene clusters are useful for the production
CC of biologically active macroclides. The macrolide biosynthetic proteins
CC are useful for synthesis of methymycin, pikromycin, neomethymycin and
CC narbomycin. The alternative termination of polyketide synthases may be
CC useful to prepare novel antibiotics and polyhydroxyalkanoate (PHA)
CC monomers. The compounds produced by the recombinant host cells are useful
CC as biopolymers, e.g., in packaging or biomedical applications, to
CC engineer PHA monomer synthases or to prepare biologically active agents,
CC such as chemotherapeutics, immunosuppressants, agents to treat asthma,
CC chronic obstructive pulmonary disease as well as other diseases involving
CC respiratory inflammation, cholesterol-lowering agents or macrolide-based
CC antibiotics which are active against a variety of organisms, e.g.,
CC bacteria, including multi-drug resistant pneumococci and other
CC respiratory pathogens, as well as viral parasitic pathogens, or as crop
CC protection agents (e.g., fungicides or insecticides) via expression of
CC polyketides in plants. The present sequence represents the macrolide
CC biosynthetic gene cluster (pik) from *Streptomyces venezuelae* ATCC
CC 15439, as given in figure 31.
XX
SQ Sequence 36778 BP; 4758 A; 14303 C; 12864 G; 4851 T; 2 other;
Query Match 38.2%; Score 38.6; DB 21; Length 36778;
Best Local Similarity 61.4%; Pred. No. 0.14;
Matches 62; Conservative 0; Mismatches 39; Indels 0; Gaps 0;
QY 1 GTGCAGGCCGAGATCGCGCGCTTCATGAGGGCCGCCGACGCCGCTCCGTCGAT 60
DB 32632 GTGATGCGGCGAGCGGCTCCGTCCGGTCCGACATCGCCGAAGATCCCGTCGAC 32691
QY 61 CGCGCGCTGTCGACTTGCGCTCGACTCGCTACGCGGAT 101
DB 32692 CGCCCGCTGGGAGATCGGCTTCGACTCGCTGACCGCCGT 32732
RESULT 12
AA287285
ID AA287285 standard; DNA: 37948 BP.
XX
AC AA287285;
XX
DT 05-JUN-2000 (first entry)
XX
DE S. venezuelae pik (macrolide biosynthesis) gene cluster. seq ID NO:5.
XX
KM Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;
KM neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
KM biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;
KM chronic obstructive pulmonary disease; respiratory inflammation;
KM hypercholesterolemia; crop protection agent; ds.
XX
OS Streptomyces venezuelae ATCC15439.
XX
PN W0200000620-A2.
XX
PD 06-JAN-2000.
XX
PF 25-JUN-1999; 99MO-US14398.
XX
PR 26-JUN-1998; 98US-0105537.
XX
PA (MINU) UNIV MINNESOTA.
XX
PI Sherman DH, Liu H, Xue Y, Zhao L;
XX

DR WPI: 2000-160679/14.
DR P-PSDB: AAY77180.
XX
PT Desosamine and macrolide biosynthetic gene clusters, useful for, e.g.,
PT synthesis of methymycin and pikromycin -
XX
PS Claim 13; Page 299-315; 438pp; English.
XX
XX
CC The invention relates to an isolated and purified nucleic acid segment
CC comprising a desosamine biosynthetic gene cluster, a fragment or its
CC biologically active variant, where the nucleic acid sequence is not
CC derived from the eryC gene cluster of *Saccharopolyspora erythraea* or
CC streptomycetes antibiotics. The invention also relates to a macrolide
CC biosynthetic gene cluster, or fragments thereof. The macrolide
CC biosynthetic gene cluster encodes proteins which synthesise methymycin,
CC pikromycin, neomethymycin, narbomycin or a combination of these
CC compounds. Recombinant or augmented cells comprising the desosamine
CC and/or macrolide biosynthetic gene clusters are useful for the production
CC of biologically active macroclides. The macrolide biosynthetic proteins
CC are useful for synthesis of methymycin, pikromycin, neomethymycin and
CC narbomycin. The alternative termination of polyketide synthases may be
CC useful to prepare novel antibiotics and polyhydroxyalkanoate (PHA)
CC monomers. The compounds produced by the recombinant host cells are useful
CC as biopolymers, e.g., in packaging or biomedical applications, to
CC engineer PHA monomer synthases or to prepare biologically active agents,
CC such as chemotherapeutics, immunosuppressants, agents to treat asthma,
CC chronic obstructive pulmonary disease as well as other diseases involving
CC respiratory inflammation, cholesterol-lowering agents or macrolide-based
CC antibiotics which are active against a variety of organisms, e.g.,
CC bacteria, including multi-drug resistant pneumococci and other
CC respiratory pathogens, as well as viral parasitic pathogens, or as crop
CC protection agents (e.g., fungicides or insecticides) via expression of
CC polyketides in plants. The present sequence represents the macrolide
CC biosynthetic gene cluster (pik) from *Streptomyces venezuelae* ATCC 15439.
XX
SQ Sequence 37948 BP; 4901 A; 14718 C; 13311 G; 5018 T; 0 other;
Query Match 38.2%; Score 38.6; DB 21; Length 37948;
Best Local Similarity 61.4%; Pred. No. 0.14;
Matches 62; Conservative 0; Mismatches 39; Indels 0; Gaps 0;
QY 1 GTGCAGGCCGAGATCGCGCGCTTCATGAGGGCCGCCGACGCCGCTCCGTCGAT 60
DB 35802 GTGATGCGGCGAGCGGCTCCGTCCGGTCCGACATCGCCGAAGATCCCGTCGAC 35861
QY 61 CGCGCGCTGTCGACTTGCGCTCGACTCGCTACGCGGAT 101
DB 35862 CGCCCGCTGGGAGATCGGCTTCGACTCGCTGACCGCCGT 35902
RESULT 13
AAA75633
ID AAA75633 standard; DNA: 38506 BP.
XX
AC AAA75633;
XX
DT 22-JAN-2001 (first entry)
XX
DE Nucleotide sequence of the insert DNA in cosmid pKOS023-27.
XX
KM Narbonolide synthase; polyketide synthase gene; narbonolide polyketide;
KM antibiotic; C12-hydroxylase; pik; desosamine biosynthesis;
KM desosaminyl transferase enzyme; ketolide; beta-glucosidase enzyme;
KM pikromycin biosynthesis; ss.
XX
OS Streptomyces venezuelae.
XX
PN US6117659-A.
XX
PD 12-SEP-2000.
XX
PF 27-MAY-1999; 99US-0320878.
XX

PR	28-MAY-1998;	98US-0087080.		
PR	22-SEP-1998;	98US--0100880.		
PR	08-FEB-1999;	99US-0119139.		
PR	20-MAY-1999;	99US-0134990.		
PR	30-APR-1997;	97US-0846247.		
PR	06-MAY-1998;	98US-0073538.		
PR	28-AUG-1998;	98US-0141908.		
XX	(KOSA-) KOSAN BIOSCIENCES INC.			
PA				
PI	Ashley G, Betlach MC, Betlach M, Tang L, McDaniel R;			
XX				
DR	WPI: 2000-610844/58.			
XX				
PT	New recombinant pick hydroxylase gene of Streptomyces venezuelae useful			
PT	for converting ketolides to antibiotics and as antibiotics and			
PT	intermediates in the synthesis of compounds with pharmaceutical value			
PT	-			
PS	Disclosure: Columns 15-32; 117p; English.			
XX				
CC	The present sequence is used to produce the recombinant DNA compounds			
CC	of the invention. The specification describes a recombinant DNA compound			
CC	expressing recombinant polyketide synthase genes in host cells for the			
CC	production of narbonolide, narbonolide synthase derivatives and polyketides that			
CC	are useful as antibiotics and as intermediates in the synthesis of			
CC	compounds with pharmaceutical value. The DNA compounds may also encode			
CC	a C12-hydroxylase (p1ck), desosamine biosynthesis and desosaminyl			
CC	transferase enzymes (useful for conversion of ketolides to antibiotics),			
CC	and the beta-glucosidase enzyme (involved in plicromycin biosynthesis).			
CC	These compounds are also useful for increasing the antibiotic activity			
CC	of a compound relative to the unhydroxylated compound. The recombinant			
CC	host cells are useful as genetic systems that allow rapid engineering			
CC	of the narbonolide polyketide synthase. These would be valuable for			
CC	creating novel ketolide analogs for pharmaceutical applications.			
XX				
SO	Sequence 38506 BP; 4914 A; 15118 C; 13444 G; 5030 T; 0 other:			
Query Match				
Best Local Similarity 38.2%; Score 38.6; DB 21; Length 38506;				
Matches 62; Conservative 0; Mismatches 39; Indels 0; Gaps 0;				
OY	1 GGGCAGGCCGGAATCGCGCGCTGCTTCATGGGCGCCGCGAGCGCCCTGCCCTGCAT 60			
DB	32774 GGGATCGGAGCGAGCGCGCTTCCTGCTCCGGTCCGACTCGCCCGAAGAGGTCCCGTCGAC 32833			
OY	61 CGCGCGCTGTGCGACTTGGGCGCTGCAGTCGCTCAGCGGCGGT 101			
DB	32834 CGCCGCTGGCGGAGATCGGCTTCGACTCGCTGACCGCCGT 32874			
RESULT 14				
AAZ56001	AAZ56001 standard; DNA: 38506 BP.			
XX	AAZ56001;			
XX				
DT	23-MAR-2000 (first entry)			
XX				
DE	Recombinant cosmid pKOS023-27 containing S. venezuelae PKS genes.			
XX				
KW	Narbonolide polyketide synthase; PKS; cosmid pKOS023-27; ketolide;			
KW	antibiotic production; narbomycin; plicromycin; ds.			
XX				
OS	Streptomyces venezuelae.			
XX				
FT	Key			
FT	Location/Qualifiers			
FT	CDS			
FT	70..13725			
FT	/tag= a			
FT	/product= PICAT			
FT	/note= "Narbonolide synthase subunit 1"			
FT	13850..25049			
FT	/*tag= b			
FT	CDS			

FT	FT	/product= PICAIII
FT	/note="Narbonolide synthase subunit 2"	
FT	25133..29821	
FT	/tag= c	
FT	/product= PICAIII	
FT	/note="Narbonolide synthase subunit 3"	
FT	29924..33964	
FT	/tag= d	
FT	/product= PICAIIV	
FT	/note="Narbonolide synthase subunit 4"	
FT	33961..34806	
FT	/tag= e	
FT	/product= PICB	
FT	/note="Contains typeII thioesterase domain"	
FT	34863..36011	
FT	/tag= f	
FT	/product= PICCII	
FT	/note="4-keto-6-deoxyglucose isomerase"	
FT	36159..37439	
FT	/tag= g	
FT	/product= PICCIII	
FT	/note="Desosaminyl transferase"	
FT	37529..38242	
FT	/tag= h	
FT	/product= PICCIV	
FT	/note="3-amino dimethyltransferase"	
XX	MO9961599-A2.	
XX	02-DEC-1999.	
XX	27-MAY-1999;	99MO-00511814.
XX	28-MAY-1998;	98US-0087080.
XX	28-AUG-1998;	98US-0141908.
XX	22-SEP-1998;	98US-0100808.
XX	08-FEB-1999;	99US-0119139.
XX	(KOSA-) KOSAN BIOSCIENCES INC.	
XX	Ashtley G, Betlach MC, Betlach M, McDaniel R, Tang L;	
XX	WPI: 2000-072618/06.	
XX	P-PSDB: AAY67201, AAY67202, AAY67203, AAY67204, AAY67205, AAY67207,	
XX	AAY67208, AAY67211.	
XX	New recombinant DNA encoding a domain of narbonolide polyketide	
XX	synthase, for production of ketolide antibiotics -	
XX	Example 2; Page 16-27; 98pp; English.	
XX	This is the recombinant cosmid pKOS023-27 DNA sequence which contains a	
XX	Streptomyces venezuelae DNA insert. The cosmid contains open reading	
XX	frames which encode the various modules of the narbonolide polyketide	
XX	synthase (PKS). The invention relates to recombinant DNA containing a	
XX	coding sequence for a narbonolide PKS. Polyketides are compounds	
XX	synthesised from 2-carbon units through a series of condensations and	
XX	subsequent modifications. Modular PKSs are responsible for the production	
XX	of many antibiotics including plicomycin. The narbonolide PKS consists of	
XX	a loading module, six extender modules, and two thioester domains. Four	
XX	proteins make up the narbonolide PKS (PICAI, PICAI, PICAI, PICAI and PICAIIV).	
XX	PICAI includes the loading module and extender modules 1 and 2, PICAI	
XX	includes extender modules 3 and 4, PICAI includes extender module 5 and	
XX	PICAI includes extender module 6 and a type II thioesterase domain. The	
XX	second type II thioesterase domain is found on the PICB protein. The	
XX	nucleotide sequences encoding all of these proteins can be isolated in	
XX	recombinant form from the recombinant cosmid pKOS023-27. Narbonolide is	
XX	desosaminylated in S. venezuelae to yield narbornycin, the desosaminyl	
XX	transferase enzyme is required for this conversion, and the desosamine	
XX	biosynthetic genes are also found in cosmid pKOS023-27. The recombinant	
XX	DNA of the invention is used to express, in transformed cells,	
XX	narbonolide (or its derivatives) or other ketolides (particularly	
XX	hybrids), which may then be converted (e.g. by other enzymes	

CC recombinantly expressed in the same hosts) to polyketide antibiotics or
CC their intermediates. The antibiotics are useful in human or veterinary
CC medicine.

SO Sequence 38506 BP; 4914 A; 15118 C; 13444 G; 5030 T; 0 other;

Query Match 38.2%; Score 38.6; DB 21; Length 38506;

Best Local Similarity 61.4%; Pred. No. 0.14;

Matches 62; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

OY 1 GTGCAGCGCGAGATCGCGCGCTTTTCATGGGGCGCGCGCGCGCTGCCGTGCAT 60

DB 32774 GTGATGCGGAGCGCGCTCCGCTCCGCTGCGACTCGCCGAGAGGTCGCCGTGCAC 32833

OY 61 CGGCGCGCTGTCGACTTGGGCGCTCGACTCGCTCAGCGCGGT 101

DB 32834 CGCGCGCTCGGAGATCGGCTTCGACTGACCGCGCGT 32874

RESULT 15

AAA09469 ID AAA09469 standard; DNA; 50937 BP.

AC AAA09469;

DT 29-AUG-2000 (first entry)

DE Streptococcus oleandomycin gene cluster.

KW oleandomycin; oleandrolide; polyketide synthase; oleAI; oleAII; oleAIII;
KW PKS; type I; 8,8a-deoxyoleandrolide synthase; modular; ketosynthase;

KW acyl-transferase; acyl carrier protein; inactivated; polyketide;
KW macrolactone; antibiotic; molitide; erythromycin; ss.

OS Streptococcus antibioticus.

XX Key Location/Qualifiers

FT CDS 152..1426

FT /*tag- a

FT complement (1528..2637)

FT /*tag- b

FT complement (2658..4967)

FT /*tag- c

FT complement (4968..50937)

FT /*tag- d

FT complement (50938..50937)

FT /*tag- e

FT complement (50938..50937)

FT /*tag- f

FT complement (50938..50937)

FT /*tag- g

FT complement (50938..50937)

FT /*tag- h

FT complement (50938..50937)

FT /*tag- i

FT complement (50938..50937)

FT /*tag- j

FT complement (50938..50937)

FT CDS 46491..47714

FT /*tag- 1

FT complement (47808..49517)

FT /*tag- m

FT complement (49518..50937)

FT /*tag- n

FT complement (50938..50937)

FT /*tag- o

FT complement (50938..50937)

FT /*tag- p

FT complement (50938..50937)

FT /*tag- q

FT complement (50938..50937)

FT /*tag- r

FT complement (50938..50937)

FT /*tag- s

FT complement (50938..50937)

FT /*tag- t

FT complement (50938..50937)

FT /*tag- u

FT complement (50938..50937)

FT /*tag- v

FT complement (50938..50937)

FT /*tag- w

FT complement (50938..50937)

FT /*tag- x

FT complement (50938..50937)

FT /*tag- y

FT complement (50938..50937)

FT /*tag- z

FT complement (50938..50937)

FT /*tag- aa

FT complement (50938..50937)

FT /*tag- ab

FT complement (50938..50937)

FT /*tag- ac

FT complement (50938..50937)

Query Match 38.2%; Score 38.6; DB 21; Length 50937;

Best Local Similarity 61.4%; Pred. No. 0.14;

Matches 62; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

OY 1 GTGCAGCGCGAGATCGCGCGCTTTTCATGGGGCGCGCGCGCGCTGCCGTGCAT 60

DB 39231 GTCCGATCCGAGCGCGCTGCTGCGGCGCGCGCGCGCGCGCGCGCGCGAG 39290

OY 61 CGGCGCGCTGTCGACTTGGGCGCTCGACTCGCTCAGCGCGGT 101

DB 39291 CGGCGCGCTCAGGAGCTCGGCTTCGACTGACCGCGGT 39331

Fri Nov 8 09:18:40 2002

us-09-724-876-2_copy_31170_31270.rng

Page 13

Search completed: November 6, 2002, 12:16:36
Job time : 271 secs

Om nucleic - nucleic search, using sw model

Run on: November 6, 2002, 11:51:49 ; Search time 443.625 Seconds

(without alignments)
6625.829 Million cell updates/sec

Title: US-09-724-876-2_COPY_31170_31270
Perfect score: 101

Sequence: 1 gtgcagcgcgagatcgcg...ctgcactcgctcacgcggt 101

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

```
Minimum DB seq length: 0
```

Maximum DB seq length: 200000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

1:	gd_da.*
2:	gd_hvg.*
3:	gd_in.*
4:	gd_com.*
5:	gd_ov.*
6:	gd_pat.*
7:	gd_ph.*
8:	gd_pl.*
9:	gd_pr.*
10:	gd_ro.*
11:	gd_scs.*
12:	gd_sy.*
13:	gd_un.*
14:	gd_vl.*
15:	em_ba.*
16:	em_fun.*
17:	em_hum.*
18:	em_in.*
19:	em_mn.*
20:	em_com.*
21:	em_or.*
22:	em_ov.*
23:	em_pat.*
24:	em_pl.*
25:	em_pl.*
26:	em_ro.*
27:	em_scs.*
28:	em_un.*
29:	em_vl.*
30:	em_hvg_hum.*
31:	em_hvg_inv.*
32:	em_hvg_other.*
33:	em_hvg_mn.*
34:	em_hvg_pln.*
35:	em_hvg_pod.*
36:	em_hvg_mam.*
37:	em_hvg_vrt.*
38:	em_sy.*
39:	em_hvg_hum.*
40:	em_hvg_mus.*
41:	em_hvg_others.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result		Query				ID	Description
No.	Score	Match	Length	DB			
1	101	100.0	58733	1	AF217189	AF217189 Sorangium	
2	101	100.0	71989	6	AR172664	AR172664 Sequence	
3	99.4	98.4	68750	6	AF193043	AF193043 Sorangium	
4	99.4	98.4	68750	6	AR193029	AR193029 Sequence	
5	99.4	98.4	68750	6	AR199551	AR199551 Sequence	
6	99.4	98.4	68750	6	AR199559	AR199559 Sequence	
7	99.4	98.4	68750	6	AR199567	AR199567 Sequence	
8	99.4	98.4	68750	6	AR201097	AR201097 Sequence	
9	99.4	98.4	68750	6	AR208671	AR208671 Sequence	
10	48.2	47.7	53784	1	AMM223012	AMM223012 Amycolato	
11	48.2	47.7	53789	6	A69720	A69720 Sequence 3	
12	48.2	47.7	80445	1	AF040570	AF040570 Amycolato	
13	41.8	41.4	82746	1	AF453501	AF453501 Actinosyn	
14	40.4	40.0	16767	6	AX089464	AX089464 Sequence	
15	40.4	40.0	25360	6	AX089418	AX089418 Sequence	
16	40.4	40.0	29736	6	AX089421	AX089421 Sequence	
17	40.4	40.0	80161	1	AY007564	AY007564 Saccharop	
18	40.4	40.0	80161	6	AR165018	AR165018 Sequence	
19	40.2	39.8	84985	1	SMA278573	AJ278573 Streptomy	
20	38.6	38.2	11480	1	STMPS30RF	L09654 Streptomyc	
21	38.6	38.2	24225	1	SC2C4	AL0512902 Streptomyc	
22	38.6	38.2	27541	6	AF079136	AF079136 Sequence	
23	38.6	38.2	37948	6	AF079138	AF079138 Streptomyc	
24	38.6	38.2	50937	6	AR159871	AR159871 Sequence	
25	38.6	38.2	78210	1	AB070949	AB070949 Streptomyc	
26	38.6	38.2	113193	1	AF357202	AF357202 Streptomyc	
27	38.6	38.2	123580	1	AF263912	AF263912 Streptomyc	
28	38.6	38.2	125401	6	AX211739	AX211739 Sequence	
29	37	36.6	9513	6	AX089460	AX089460 Sequence	
30	37	36.6	14775	6	AX089462	AX089462 Sequence	
31	37	36.6	20394	1	SNM132222	AJ132222 Streptomyc	
32	37	36.6	20394	1	AX067996	AX067996 Sequence	
33	37	36.6	43280	1	SF078289	U78289 Streptomyc	
34	37	36.6	50000	6	AX089417	AX089417 Sequence	
35	37	36.6	50000	6	AX089420	AX089420 Sequence	
36	37	36.6	104326	1	AB070940	AB070940 Streptomyc	
37	35.8	35.4	65140	6	AX211705	AX211705 Sequence	
38	35.4	35.0	1454	6	AR173224	AR173224 Sequence	
39	35.4	35.0	5676	6	A69718	A69718 Sequence 1	
40	35.4	35.0	15970	1	AE007061	AE007061 Mycobacte	
41	35.4	35.0	27522	1	AB070942	AB070942 Streptomyc	
42	35.4	35.0	30000	6	AX250261	AX250261 Sequence	
43	35.4	35.0	30000	6	AX250263	AX250263 Sequence	
44	35.4	35.0	41097	1	AF016585	AF016585 Streptomyc	
45	35.4	35.0	53450	1	MTF018	AL021899 Mycobacte	

ALIGNMENTS

RESULT 1	AF217189	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE
	AF217189	58733 bp	DNA	linear	BCT 09-JUN-2000				
	Sorangium cellulosum putative transposase gene, partial cds; putative transposase gene, complete cds; epothione biosynthesis gene cluster, complete sequence; putative membrane protein gene, complete cds.								
	AF217189								
	AF217189.1	GI:7453554							
	Polyangium cellulosum . Polyangium cellulosum . Bacteria; Proteobacteria; delta subdivision; Myxobacteria; Myxococcales; Sorangineae; Polyangiaceae; Polyangium . (bases 1 to 58733)								

Best Local Similarity 99.0%; Pred. No. 7.5e-12;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GTGCAGGCCGAGATCGCCGCGTCTTCATGAGGCGCCGAGCGCCGCTGCAT 60
|||||
Db 36782 GTGCAGGCCGAGATCGCCGCGTCTTCATGAGGCGCCGAGCGCCGCTGCAT 36841

Qy 61 CGGCCGCTGTGCGACTTGGGCGCTCGACTCGCTCAGCGGGT 101
|||||
Db 36842 CGGCCGCTGTGCGACTTGGGCGCTCGACTCGCTCAGCGGGT 36882

RESULT 4
ARI93029 68750 bp DNA linear PAT 20-APR-2002
LOCUS
DEFINITION Sequence 1 from patent US 6346404.
ACCESSION ARI93029
VERSION ARI93029.1 GI:20238994
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE Unclassified.
AUTHORS 1 (bases 1 to 68750)
Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and
Gorlach,J.
TITLE Genes for the biosynthesis of epoethlones
JOURNAL Patent: US 6346404-A 1 12-FEB-2002;
FEATURES
source 1. .68750
/organism="unknown"

BASE COUNT 9596 a 22456 c 25539 g 11159 t
ORIGIN

Query Match 98.4%; Score 99.4; DB 6; Length 68750;
Best Local Similarity 99.0%; Pred. No. 7.5e-12;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GTGCAGGCCGAGATCGCCGCGTCTTCATGAGGCGCCGAGCGCCGCTGCAT 60
|||||
Db 36782 GTGCAGGCCGAGATCGCCGCGTCTTCATGAGGCGCCGAGCGCCGCTGCAT 36841

Qy 61 CGGCCGCTGTGCGACTTGGGCGCTCGACTCGCTCAGCGGGT 101
|||||
Db 36842 CGGCCGCTGTGCGACTTGGGCGCTCGACTCGCTCAGCGGGT 36882

RESULT 5
ARI99551 68750 bp DNA linear PAT 20-APR-2002
LOCUS
DEFINITION Sequence 1 from patent US 6355457.
ACCESSION ARI99551
VERSION ARI99551.1 GI:20249625
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE Unclassified.
AUTHORS 1 (bases 1 to 68750)
Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and
Gorlach,J.
TITLE Genes for the biosynthesis of epoethlones
JOURNAL Patent: US 6355457-A 1 12-MAR-2002;
FEATURES
source 1. .68750
/organism="unknown"

BASE COUNT 9596 a 22456 c 25539 g 11159 t
ORIGIN

Query Match 98.4%; Score 99.4; DB 6; Length 68750;
Best Local Similarity 99.0%; Pred. No. 7.5e-12;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GTGCAGGCCGAGATCGCCGCGTCTTCATGAGGCGCCGAGCGCCGCTGCAT 60
|||||
Db 36782 GTGCAGGCCGAGATCGCCGCGTCTTCATGAGGCGCCGAGCGCCGCTGCAT 36841

Db 36782 GTGCAGGCCGAGATCGCCGCGTCTTCATGAGGCGCCGAGCGCCGCTGCAT 36841

Qy 61 CGGCCGCTGTGCGACTTGGGCGCTCGACTCGCTCAGCGGGT 101
|||||
Db 36842 CGGCCGCTGTGCGACTTGGGCGCTCGACTCGCTCAGCGGGT 36882

RESULT 6
ARI99559 68750 bp DNA linear PAT 20-APR-2002
LOCUS
DEFINITION Sequence 1 from patent US 6355458.
ACCESSION ARI99559
VERSION ARI99559.1 GI:20249633
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE Unclassified.
AUTHORS 1 (bases 1 to 68750)
Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and
Gorlach,J.
TITLE Genes for the biosynthesis of epoethlones
JOURNAL Patent: US 6355458-A 1 12-MAR-2002;
FEATURES
source 1. .68750
/organism="unknown"

BASE COUNT 9596 a 22456 c 25539 g 11159 t
ORIGIN

Query Match 98.4%; Score 99.4; DB 6; Length 68750;
Best Local Similarity 99.0%; Pred. No. 7.5e-12;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GTGCAGGCCGAGATCGCCGCGTCTTCATGAGGCGCCGAGCGCCGCTGCAT 60
|||||
Db 36782 GTGCAGGCCGAGATCGCCGCGTCTTCATGAGGCGCCGAGCGCCGCTGCAT 36841

Qy 61 CGGCCGCTGTGCGACTTGGGCGCTCGACTCGCTCAGCGGGT 101
|||||
Db 36842 CGGCCGCTGTGCGACTTGGGCGCTCGACTCGCTCAGCGGGT 36882

RESULT 7
ARI99567 68750 bp DNA linear PAT 20-APR-2002
LOCUS
DEFINITION Sequence 1 from patent US 6355459.
ACCESSION ARI99567
VERSION ARI99567.1 GI:20249641
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE Unclassified.
AUTHORS 1 (bases 1 to 68750)
Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and
Gorlach,J.
TITLE Genes for the biosynthesis of epoethlones
JOURNAL Patent: US 6355459-A 1 12-MAR-2002;
FEATURES
source 1. .68750
/organism="unknown"

BASE COUNT 9596 a 22456 c 25539 g 11159 t
ORIGIN

Query Match 98.4%; Score 99.4; DB 6; Length 68750;
Best Local Similarity 99.0%; Pred. No. 7.5e-12;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GTGCAGGCCGAGATCGCCGCGTCTTCATGAGGCGCCGAGCGCCGCTGCAT 60
|||||
Db 36782 GTGCAGGCCGAGATCGCCGCGTCTTCATGAGGCGCCGAGCGCCGCTGCAT 36841

Qy 61 CGGCCGCTGTGCGACTTGGGCGCTCGACTCGCTCAGCGGGT 101
|||||
Db 36842 CGGCCGCTGTGCGACTTGGGCGCTCGACTCGCTCAGCGGGT 36882

RESULT 8
AR201097
LOCUS AR201097 68750 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 1 from patent US 6358719.
ACCESSION AR201097
VERSION AR201097.1 GI:20251985
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and Gorlach,J.
TITLE Genes for the biosynthesis of epothilones
JOURNAL Patent: US 6358719-A 1 19-MAR-2002;
FEATURES
source Location/Qualifiers
1..68750
/organism="unknown"
BASE COUNT 9596 a 22456 c 25539 g 11159 t
ORIGIN
Query Match 98.4%; Score 99.4; DB 6; Length 68750;
Best Local Similarity 99.0%; Pred. No. 7.5e-12;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 GTGAGGCGGAGATCGGCGCGCTTTCATGGGGCGCGGAGCGCGCGTCCGTCGAT 60
|||||
Db 36782 GTGAGGCGGAGATCGGCGCGCTTTCATGGAGCGCGCGGAGCGCGCGTCCGTCGAT 36641
|||||
OY 61 CGGCGCGTGTGCGACTGGGCTTCGACTCGCTCAGCGCGGT 101
|||||
Db 36842 CGGCGCGTGTGCGACTGGGCTTCGACTCGCTCAGCGCGGT 36882
|||||
RESULT 9
AR208671
LOCUS AR208671 68750 bp DNA linear PAT 20-JUN-2002
DEFINITION Sequence 1 from patent US 6383787.
ACCESSION AR208671
VERSION AR208671.1 GI:21509886
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and Gorlach,J.
TITLE Genes for the biosynthesis of epothilones
JOURNAL Patent: US 6383787-A 1 07-MAY-2002;
FEATURES
source Location/Qualifiers
1..68750
/organism="unknown"
BASE COUNT 9596 a 22456 c 25539 g 11159 t
ORIGIN
Query Match 98.4%; Score 99.4; DB 6; Length 68750;
Best Local Similarity 99.0%; Pred. No. 7.5e-12;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 GTGAGGCGGAGATCGGCGCGCTTTCATGGGGCGCGGAGCGCGCGTCCGTCGAT 60
|||||
Db 36782 GTGAGGCGGAGATCGGCGCGCTTTCATGGAGCGCGCGGAGCGCGCGTCCGTCGAT 36641
|||||
OY 61 CGGCGCGTGTGCGACTGGGCTTCGACTCGCTCAGCGCGGT 101
|||||
Db 36842 CGGCGCGTGTGCGACTGGGCTTCGACTCGCTCAGCGCGGT 36882
|||||
RESULT 10
AMM223012
LOCUS AMM223012 53784 bp DNA linear BCT 09-FEB-1998

DEFINITION Amycolatopsis mediterranei genes encoding rifamycin polyketide synthases, ORFs 1 to 5.
ACCESSION AJ223012
VERSION AJ223012.1 GI:2764760
KEYWORDS ORF1; ORF2; ORF3; ORF4; ORF5; polyketide synthase; rifamycin.
SOURCE Amycolatopsis mediterranei.
ORGANISM Amycolatopsis mediterranei
Bacteria; Actinobacteridae; Actinomycetales; Pseudonocardiaceae; Amycolatopsis.
REFERENCE 1 (bases 1 to 53784)
AUTHORS Schupp,T.
TITLE Direct Submission
JOURNAL Submitted (12-DEC-1997) Schupp T., Core Technology, Novartis Pharma AG, CH 4002 Basel, Postfach, Novartis K-681.344, 4002 Basel, SWITZERLAND
REFERENCE 2 (bases 1 to 53784)
AUTHORS Schupp,T., Toupet,C., Engel,N. and Goff,S.
TITLE Cloning and sequence analysis of the putative rifamycin polyketide synthase gene cluster from Amycolatopsis mediterranei
JOURNAL Unpublished
FEATURES
source Location/Qualifiers
1..53784
/organism="Amycolatopsis mediterranei"
/strain="LBG A3136"
/db_xref="taxon:33910"
1336..15543
/function="CoA ligase, acylcarrier protein, b-ketocacyl-ACP-synthase, acyltransferase, ketoreductase"
/note="ORF1. PKS modules 1-3"
/codon_start=1
/transl_table=11
/product="rifamycin polyketide synthase, type 1"
/protein_id="CAA11035.1"
/db_xref="GI:2764761"
/db_xref="SPTREMBL:O54666"
/translation="MRTDLIKPLVLALENATRPGRKPAFADHRTVYTGDEATRR LAGHLGLGVRRHGRDVAICLGNRVSTVESYFAILAGAGVPLNGSATLEHPLTD SGATVVTVDAAQAARLRLAPHELVTDGDDVEGHSYDELALSPAEPAADLEDE PAMFETSGTGRPKGVSTORNLMSVASCYVPPGSLDORVLPMLPLFSLSHA CYLSATVVGASVRIADGSSADVMRLIESSSTPLAGVPTTTHHIVRAARVSCFAPS LRIAGLAGAVIAGAGLSREPEPTGYPDLIDAYSTECGATITNNPDAARVESCCLAV PGVDVAVDPDPTGLDVPAGEBEVWVSGVNWLGHSNPEATAAARDWFRITGLAR RDDAGFTTCGRALKELIRGANGIHGEVEALRVTDGADAAGVGPMDTLGEPVAV YVIGPFTGCDPALIEKCREQLSAVKVPDRIELVHAIPETAGKTRGILDEPAQLR VATHEEOSRADESVAAALRLARLSGLDERAQCLEDLDTVROADYVIGQVDPGRAP FDIQFTSLAIVELRNRLREHNGIMLPASAVDHPPALARVARELIGLIGQVAAPV VVAADPGEPIAIVGMAARLPGVYASPEDIMRLVARYVAVSFEPDRGMDLSLTDPD RERAGTSTVGCGFLHDAGEFDGFEFISPRVAVADPQQRLLLETSMALENACVDP IALKTDTGVFSGLMGQVGSAGVAPBELGFPYTTGVAASSVAGRSVYLGLEGAVTY DPAQSSSLVAMHLAAQALROGECSMALAGGVVMTPGSFVFSRQRLADPGRCKAF AAADGTGWESEGVVLERLSVAREGRHILAVRGSAVNDGASNGLTPANGLSQO RYIIRLALAAVGLAPSDVDVYEAHGGTTLGDPLEAOALATYGOBRKOPIMPTKSKN IGHAAQAAVAGVATIKVQALREHETLPRTIHLVCKEPLTEVDMWSGALIELTEARAPRNG RPRKGVSSFGVSGTNALILTEAPREPEVYAPBELPVPLVVSANSTESLSQARLA SLLEGDSVLTEVAGALVSRRVAVDERAVVAASREAVTGLALMTAGSGTGRKVVW PFGQGTOMAGMGRLELAEPSVFAERIAECALAPWIMSDLVLRGDLGRVDVLQ PACFAGMVGLAAWESVGVPRDPAVGVGSHQGEITAAACVSGALSLEDAAKVALASQAIA AELSGKGASVALGEDVDVSVLVGVGEVAAVNGSSVYIADDAHALATTEILISGEG IYRRVAVDYASHTRHVEDIRDTLAETFAETISAOAPAYFVSTVSEVVRADGVDCG YTRRLNRNOYRFGAATYALLLEDGHTFVGVESHPTVVOPLSLTDTAIGTLRREDCGL RRLASMGELFVRGIDVMTAVPAPAGVMDLPTVFAEHHHWLEPAEPAASADPLLGT VVSTPGSDRLTVAAQMSRRAPWADGILPNAALVEA1LRIDLAGTAPVGLVVDAP VVLPGRGSEVOLIIVGPEQGRRRPEVRSRADPMTWRAHAGTLPAPAAVPEPAAAG GATQDVTYVAGLDADRYGHIHPALLDAAVETVGGDLDLSVMTGVSLLASGATVAVVTP TATGRLTPDRAGQPVLTVESVSGTPEVFAEGTDTDLAFDVMPEITLIPAEARDELPEY ATSAENTLSALDAMLADPAETRLAVVGTGCTEGPAAALMGVLVRSQSHPEGRITVIALD DDPAYLPAVAVASGEPOVVRNCAVSPRLTRVTPQDARLAPDPESTVLTIGTGFLGA LTAHRLVTAHGVNRHLVIVSRREAEVLEIELTALCAVSAIACADVADBAOQAEVAFRAI PAEHPLTAVIHTAGVLDGCVTELTPDRLATVRRKVDARLDELTEADLAAGVLF SAAGVIGNPGOAGYAANAELDALARONSIDLPAVSAIANGVYATVGTGMEHIGDAD LRRNORIGMSGLPADEGMALDLDAIATGTTLYAAKFDVAALBATKAGGPPVPLRLGL APVPRRAAKTASLTERLAGLAETSDAAALDLVVRHAAEVYGHSGASVSHSGRFKID AGFDSLTAIVELRNRLAAATGTLTSPAMIFDYKPPALADHLRAKLFGSANRPAEIGT

KEYWORDS	unidentified.
SOURCE	unidentified.
ORGANISM	unclassified.
REFERENCE	1 (bases 1 to 53789)
AUTHORS	Schupp,T., Roupet,C. and Engel,N.
TITLE	RIFAMYCIN BIOSYNTHESIS GENE CLUSTER
JOURNAL	Patent: WO 9807868-A 3 26-FEB-1998;
FEATURES	CIBA GEIGY AG (CH)
source	Location/Qualifiers
	1..53789
	/organism="unidentified"
	/db_xref="taxon:32644"
BASE COUNT	6707 a 19183 c 20504 g 7395 t
ORIGIN	
Query Match	47.7%; Score 48.2; DB 6; Length 53789;
Best Local Similarity	67.3%; Pred. No.0.34; Mismatches 33; Indels 0; Gaps 0;
Matches	68; Conservative 0; Mismatches 33; Indels 0; Gaps 0;
OY	1 GTGCAGCCGAGATCGCGCGCTGCTTCATGAGGGCGCGGAGCGCCGTGCCCTGCAT 60
Db	25279 GTCCGCGCGAGAGGTGCGGCGCTGCTCGGCGACGCGGACGAGCGGCGCTCGCTGCAC 25338
OY	61 CGGCGCGTGTGCGACTTGGGCGCTGCACATCGCTGCACGCGCGGT 101
Db	25339 ACGGCTTCATAGACGCGCGCTTCGATCGCTGCACCGCGGT 25379
RESULT 12	
LOCUS	AF040570 90445 bp DNA linear BCT 16-APR-2001
DEFINITION	Amycolatopsis mediterranei rifamycin biosynthetic gene cluster.
ACCESSION	AF040570 AF040571
VERSION	AF040570.2 GI:12331604
KEYWORDS	
SOURCE	Amycolatopsis mediterranei.
ORGANISM	Amycolatopsis mediterranei
	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
	Actinomycetales; Pseudonocardineae; Pseudonocardaceae;
	Amycolatopsis.
REFERENCE	1 (bases 1 to 90445)
AUTHORS	August,P.R., Tang,L., Yoon,Y.J., Ning,S., Mueller,R., Yu,T.W.,
	Taylor,M., Hoffmann,D., Kim,C.G., Zhang,X., Hutchinson,C.R. and
	Floss,H.G.
TITLE	Biosynthesis of the ansamycin antibiotic rifamycin: deductions from
	the molecular analysis of the rif biosynthetic gene cluster of
JOURNAL	Amycolatopsis mediterranei S699
MEDLINE	Chem. Biol. 5 (2), 69-79 (1998)
PUBMED	98174059
REFERENCE	9512878
AUTHORS	2 (bases 1 to 90445)
TITLE	Kim,C.G., Yu,T.W., Fryhle,C.B., Handa,S. and Floss,H.G.
	3-amino-5-hydroxybenzoic acid synthase, the terminal enzyme in the
	formation of the precursor of mcfN units in rifamycin and related
	antibiotics
JOURNAL	J. Biol. Chem. 273 (11), 6030-6040 (1998)
MEDLINE	98165773
PUBMED	9497318
REFERENCE	3 (bases 1 to 90445)
AUTHORS	Yu,T.W., Muller,R., Muller,M., Zhang,X., Draeger,G., Kim,C.G.,
	Leisner,E. and Floss,H.G.
TITLE	Mutational analysis and reconstituted expression of the
	biosynthetic genes involved in the formation of
	3-amino-5-hydroxybenzoic acid, the starter unit of rifamycin
	biosynthesis in Amycolatopsis mediterranei S699
JOURNAL	J. Biol. Chem. 276 (16), 12546-12555 (2001)
MEDLINE	11278540
PUBMED	12120176
REFERENCE	4 (bases 1 to 90445)
AUTHORS	August,P.R., Tang,L., Yoon,Y.J., Ning,S., Mueller,R.,
	Hutchinson,C.R. and Floss,H.G.
TITLE	Direct Submission

REFERENCE	AUTHORS	TITLE	JOURNAL	REMARK	COMMENT
5 (bases 1 to 90445)	August,P.R., Tang,Y.T., Yoon,Y.T., Ning,S., Mueller,R., Hutchinson,C.R. and Floss,H.G.	Direct Submission	Submitted (30-DEC-1997) Dept. of Chemistry, University of Washington, Box 351700, Seattle, WA 98195-1700, USA	Sequence update by Submitter	On or before Jan 22, 2001 this sequence version replaced gi:27923340, gi:2832263.
FEATURES		Location/Qualifiers			
SOURCE		1. 90445			
		/organism="Amycolatopsis mediterranei"			
		/strain="S699"			
		/db_xref="taxon:33910"			
		423. .1451			
CDS		/codon_start=1			
		/transl_table=11			
		/product="unknown"			
		/protein_id="AAC01706.1"			
		/db_xref="GI:27923310"			
		/translation="MAPTGGVGESEYVKTLEVALIGAGLIRLHLEAMLGAAV RVYSDGRSRELAEPFGAKAGSILELADGADADITCPASHEITLTAAGVAV CEKRLASAEAEIEYVTAERAGYRLTAHDYRAALYARLHELIVASGRIGEGALGRF SFSYVHPRPWTGHASARSQGLITDQLHGDALHWVGDVVRHACTYQGDATAPBGD AVAGVAVLTHGASGISQVYSRMTATRPVRAVPHVSGSGSVSDSEMPQEARVVD GGAGFAVAGPSVYEDTMRFEFATAFAGGPEPRIGAKDALAAVRIIHAASAWTGRAV ELPRGAA"			
CDS		1448. .2215			
		/codon_start=1			
		/transl_table=11			
		/product="unknown"			
		/protein_id="AAC01707.1"			
		/db_xref="GI:27923311"			
		/translation="MKVAILSSPTPOAVAGALRGLPDYEVVAAASMDAEPYROAAEAG ARUCCEPPPAKETDLMKAMIDAGDRLTFPSAPCHGAFVYVRKICADGIGELITTVL SGVATSDVGVGAAPYLLDLADAVLDGEPAGQVYAOTNTVLSRIGESAAVLVVR GSVAISDVGCRHGSATGLPAVTFIGDGSVQYDAGPOLLGERPELGEDLEALMKD FLGAGDGPDPGQALRTFRILQAAVESAHGTQVLD"			
CDS		2353. .2619			
		/codon_start=1			
		/transl_table=11			
		/product="unknown"			
		/protein_id="AAC01708.1"			
		/db_xref="GI:27923312"			
		/translation="MODRSVLDGPPKOVFAAMSVTOYKIALSGSPAPLADAHRSVDSG ALENTYLRIGAGAGIEHFVHGSEFOAVDGCIDIAVFRMSHRKIAE"			
CDS		2652. .3842			
		/note="ORF0; similar to Streptomyces Sp. Chp"			
		/codon_start=1			
		/transl_table=11			
		/product="cytochrome P450 monooxygenase"			
		/protein_id="AAC01709.1"			
		/db_xref="GI:27923313"			
		/translation="MTDAISFEVPMWDTKFRDPAPVDSLSREERPLAKMYPDGHWGM IVSSYELVREVLSDLRSHSCSEVGHFVTHQGVYIPNPLIPGMFIMDPDEHRYRKK LYSGEYVRAASRLIPRAEVAABQIEVMRAKGAADVMDPAFLVLMIGELVGLD VEERDYPVAVTLLHDAADPAEAAAEVAGKPEDEVTERRORPODDIISLWPEED LTDBELRNIVTLLFAGYEFTTEGCALATGVYRALLHHTDQLAALAREPKDLNATIELLR YLVNQYHTYRTALEDKLEGLKIKQDVTVSLPANRDPAKGCAEYLEDITDTSGL HVAFGEGHOCLOGNLIARIELRAGFTALLRAFPELRLAVADEVPLRLKKSVSFVKLL PVSFG"			
gene		3883. .18090			
		/gene="rifA"			
CDS		3883. .18090			
		/gene="rifA"			
		/note="rifA"			
		/codon_start=1			
		/transl_table=11			
		/product="polyketide synthase"			
		/protein_id="AAC01710.1"			
		/db_xref="GI:27923314"			

PAFSRYVDTDRSAATGYAOGGIGADHLLNGLSLYRHSGGNAMTQRTVPTTSAGV
ARWTVRADGEQASPNDDVYQVEAPVETSAKVHTTYAGAGSYAQPSSANPN
PERGLVHTHGDDCKADLSADALRRYREBOGITAMKCFYVLAHERSSPIAQTALDROR
OPDAARVAGYKIVLFRFAYTASSESGDAPKRRVLAHTDQLPILRANADRYAVOAGFV
GANGENYITKRNFGNAGVYKATDANKRPAVLEKLLTLPADRAVQLTRPKRKLIGGG
STRIGHNDCEFLASDTPGTYEDPATEYPLLEETRLPMGETCAVNPRTCTPAT
AELSRFHWLYANTDYODDLRAMSAOGCLAETVAKLGARFTLRDSVEPTTGTGALP
VRVNLINDGYAAVNORTVNLVRNTTGALTRPLPTADPTWLPGARAIADAQIIP
SLATGVYALLLELADPALPGRPEYSIQTANTGLMDARTGINDKQITTVS"
1825..2142
/gene="asm36"
/gene="asm36"
1825..2142
/gene="asm36"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAM54114.1"
/db_xref="GI:21449382"
/translation="MYTASAVLCARAGCGAPRPAATARMFCSTACQOAYRLRKA
GSAHSHEDDEQLREALLAVARARTLTLEATPPWSAQATALVGLARRVLELAARDAGV
SGG"
complement(2129..3307)
/gene="asm35"
/complement(2129..3307)
/gene="asm35"
/codon_start=1
/transl_table=11
/product="bicyclicomycin resistant protein"
/protein_id="AAM54113.1"
/db_xref="GI:21449381"
/translation="MAVGRADSAVGMPLPALLIYALVTGTLIEIDMSVSPSPDIAREIG
VGGPVALQVTLNLFGLGCGALCDLSDRFGRPVPLGGSAYMLVAGLACALAPTG
LLTARFVOGIGASAAVLYEVIIGDRYGPALRMVGLVNAAMSALMPCAPLGGAI
NRAGRGNYLACATTAIILVLETPETRTARTAITARSLADRYLGSRAVL
AALITPSLLPAALVFIASPRLYTAAFGVGTGAPALHLVYASAVBAFPTGITP
LLGGPAGVARMGAGAVVGVGLTLGDDGVAATGSALECGFACTVPPFSRVS
FPELAGAASSATGRLALLVTLTLALAGALSTGGPPAPAAVMAAVALVLSLSPR
"
complement(3450..4064)
/gene="asm34"
complement(3450..4064)
/gene="asm34"
/codon_start=1
/transl_table=11
/product="transcriptional regulator"
/protein_id="AAM54112.1"
/db_xref="GI:21449380"
/translation="MARNRPERAALLDAAVEYLAGEGARGLTFRVNDORAGVPACTA
SNFAGRAELITGAGERVYQRLDQAPLGETPUGPRDRSVTELMHALVDVSAFPPTG
FLALLLELRLEAARNPELRVLTARADILRFNAEHHAASQLPGDCTTVALMLMALML
ILERLTPDLTLDHQRDLVTALVDRLLAGQPEAGSPDESCK"
4151..4723
/gene="asm33"
4151..4723
/gene="asm33"
/codon_start=1
/transl_table=11
/product="dihydrofolate reductase"
/protein_id="AAM54111.1"
/db_xref="GI:21449379"
/translation="MRKLVYFIASIDGVYIAGEGHDYPTAEPVEHLRAEYPERET
PTHVRPLVGMPLDTPNRRYDTVLMGRGSYOPALDAGVGSYALRQYVVSRSIPDADP
AAVALRADPVGLVRELKREDDLVMLCGGADLAGQLPEIDELVVTYVYVAGCGPL
PRAGFAPRAFPRPDVRAFDHGVVTTYRPA"
complement(4854..5057)
/gene="asm32"
complement(4854..5057)
/gene="asm32"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAM54110.1"
/db_xref="GI:21449378"

/translation="MWITQONARIPARRSSGSRNARCAVRMPQPRSATRAPGRS
SANADSRASRGCDPRTSLNRSA"
5415..6461
/gene="asm31"
5415..6461
/gene="asm31"
/codon_start=1
/transl_table=11
/product="RNA polymerase sigma factor"
/protein_id="AAM54109.1"
/db_xref="GI:21449377"
/translation="MWIGYSADGRRRRPVRMWGVTEAFWETDEDIKIATLRDAEA
ANADSTRAYLKAIGRVRLTADQEVELARVEAGLYAEHLIDLPEAAGSRPVELR
RVYRDCQAKADHLEANLRLVYSIAKRYTGKALLDVIOEGLGLVRAVEKEDYAKG
FKESTATWMIROAITRALDOSRTIRLRYHWVEYTKIARIERDLHRNGREATAEE
VALELDTPEKVELEORHARTPYSIDOTTGAENDPSIGDLIDENAVAPLEAVSSALL
SADLRVLELTLPRENGVYRMKIGFPDGVHLTHLDEGRYGYVTRETIROIESRTAKL
QHPARSRGLEGYD"
complement(6501..9518)
/gene="asm30"
complement(6501..9518)
/gene="asm30"
/codon_start=1
/transl_table=11
/product="cytochrome p450"
/protein_id="AAM54108.1"
/db_xref="GI:21449376"
/translation="MYATGTRITDGPRLPLVGNLIDVLTSDLDTPVDPLDRCHREHG
ITALTRAGORVYFASSHETVARMCSDPSNGKAAWHRLVLEVRDPAGDGLTAKDEHNG
GAKHRLMPAFGPTAMRDHPAMLIDIAEOMLVKRRFGDHRIDVADNRDTRLTLDTIA
LCAFGARFNSFEDRAHPFVDMAVRSLVAGEBAERLPGVOFLVGRNORYDDIATM
NRADQIVARARALPAGEPDDILEMLTCADPVGERLSARVRQOLATFLAGHET
TSGLSFEAHLRAHEPVLRKAKDAVDGVLGDRVPAFEDLARLDYGVLRFTLRHP
TAPAFALADEPARELGHAIGAGEPYLVMLPTLRDPAYMRDPDVPDPRPBERDE
IPACAMPPGHGRACIGRRPALQETIYLAIVLQRFDLALDPHRLTIKOTLTKP
DSLVARTRPADPGRATVETVYRHOVFAIRHGHPLHVFTGNSGSGEGLARTIG
DGAARMGATVAPLDVAVARLPAASGVVTVSSYNGAPPDNAHFVWMLTQGGPDLSG
VYLVLYCGNLDMSATYQRPVPTLIDEMAMAAGARRLERGATDARDFGDMRWYEP
LMPILSAEGCEVEGEIGPRRVEESDAADGLDLSAVYLENRELVRGPDASGRHLE
LRPDGTSYTGDPGLSVLPONHDLVARAVARTGAERVTVESAPGLGVPGVAL
RYDELITRCVDLSAPAGAGVVARIGARCCPPEPRLEIATGTATLLELRPPSCVD
LALALELPAFTRRLYSSIAEBQAEVALVSYTGVTSGTISRPRDRAVAGIAS
PPESFRPADNTVPVVLIAAGTIGIAPFRGHLRARRALGSEPGALLTFCRPELDDL
YAEFPALDMLVEDRAYSRHPDGEVRHQRLMRQDRRELVNAGARVYLCGDATR
VGPAAVEVIGRIEPCGAMDALRAGGRVATDVE"
9664..10245
/gene="asm29"
9664..10245
/gene="asm29"
/codon_start=1
/transl_table=11
/product="transcriptional regulator"
/protein_id="AAM54107.1"
/db_xref="GI:21449375"
/translation="MSAAVRREQLVAALRVKRDGIAAATTBAICAEADMDFGAFY
CFRSKQELTYALTTLGINTSLDAAMPAISPDIPADVNTKMLAATPAQADPELV
LGEIVLTAOREPELRELPAHQREYVAKAVAVHVDRLAEAGCHDYSYDRALLAEIVVS
LQGVASWFSHRGELAKASLDQFVAFPAALVR"
10393..10857
/gene="asm28"
10393..10857
/gene="asm28"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAM54106.1"
/db_xref="GI:21449374"
/translation="MTDTYTRHRTVPQNALAVGAVFLVYIAGEVPGTLTYNDGLEF
AGHSDAALLGVFKVSAHLNLVLAEGAAGIAGSARGARAFILGGGAIYAVLML
YGVYVDEHSTANVPPLNADNNLILGLAVGMIGLGAALGRNAHRAASDD"
complement(10912..11937)
/gene="asm27"
complement(10912..11937)

Query Match	Best Local Similarity	41.4%;	Score	41.8;	DB	1;	Length	82746;	
Matches	64;	Conservative	0;	Mismatches	37;	Indels	0;	Gaps	0;
QY 1	GTGCAGCGCCGAGANTCGCGCGCTTTTCATNGGGGCGCGGAGCGCGCTCCCGTCGAT	60							
Db 28437	GTCCAGGGCGCACACCGGACCGCTGCGGGACCGGACGCGCTCCCGCTGACGCGGAA	28378							
QY 61	CGGCGCGCTGTGGGACTTGGGCTTGCATCTCGTCACGGCGGT	101							
Db 28377	CGCGCGTTCAGGAGACTGGCGCTGCACTCGCTCAGCGCGCT	28337							
RESULT 14									
AX089464									
LOCUS	AX089464	16767 bp	DNA	linear	PAT	21-MAR-2001			
DEFINITION	Sequence 49 from Patent WO0116303.								
ACCESSION	AX089464								
VERSION	AX089464.1	GI:13443725							
KEYWORDS									
SOURCE									
ORGANISM	Saccharopolyspora spinosa.								
	Saccharopolyspora spinosa								
	Bacteria; Filicutes; Actinobacteria; Actinobacteridae;								
	Actinomycetales; Pseudonocardiales; Pseudonocardaceae;								
	Saccharopolyspora								
	1 (bases 1 to 16767)								
REFERENCE	Eberz, G., Moehre, V., Froede, R., Velten, R. and Salas, J. A.								
AUTHORS	Nucleic acids which code for the enzyme activities of the spinosyn								
TITLE	biosynthesis								
JOURNAL	Patent: WO 0116303-A 49 08-MAR-2001;								
	BAYER AG (DE)								
FEATURES									
Source	Location/Qualifiers								
	1..16767								
	/organism="Saccharopolyspora spinosa"								
	/db_xref="taxon:60894"								
	1..16767								
	/note="ORF22; Polyketide synthase"								
	/codon_start=1								
	/transl_table=1								
	/protein_id="CAC34895.1"								
	/db_xref="GI:13443725"								
	/translation="MANEKLREYLKRVVLELEAHERLHELEROEHDLPIAIVSGCR								
	YRGVSTPPEELMRLLVNDGGDLIAFNPEDRCNNDELDPDPGRAGTGYVRGGGLRKY								
	ADFPAGLFGISPREAQMDPQQLLLEISWEFERAGIDPESLKTGTGYVAGLITRDY								
	YLSRGETPALEFEGYFATGNAGVSAASRVATFGLRPAVTVLTCSSSLVALALHCD								
	SLRGECDLALAGGISYMATPGAFVEFSRQALASDGRCKFPADAGSGTGWEGAGMI								
	LLLETSDBARRNGHPVLAAYVGSAINODGTSNGLAPSGPAAOQVITROALNAGSPABE								
	VLYEAHGTGTALGDPITAEALITATGANSADHPILLGSKNIGHGTAAAGVAGVIT								
	KSVLAIRHREMRPSRLHIDPSQHNDMSAGAVRLITSDVMDPLDRPRACVSPSGMG								
	TNALIIVEEVSDEPVSSTETGFAFPMLPSKTETLEAQLLELVSTHEPFGELD								
	VGSLATGRAMAERAAVYVADRDSPFIRFVEVSPHLLTAGIOETLEADAGVVVGSLS								
	GQGHMOGMAEILSESSPVEFRKLAECGAATAPYVMSLLGLVLRGDDPAPLADDDVIT								
	OLAEFAMVWSLAELMRSCGVPAVVGASOELIAAVAGALSLTDVARIITAAKCDAV								
	SALNKGGMALIAIPESAVYKRIAGLEITPAAYANGSPITVSGSEPALEIOTELTA								
	ENVTORRGYIDYASHSPQIAOVORSLDRGLRGEVSEPAELAFYSTYGERIDRTERLMA								
	DYWTQNRKQPVRFQOIVARNAQDQYRFVEVSPHLLTAGIOETLEADAGVVVGSLS								
	RRGSGSRMRVLTSLAEQVRGLPVMWQDVELNTGARVPLEPTYFQFORVWLESAED								
	AGDGSGVLLSAEHPDLGAATVTLADAGGFLTGKLSYKTOPMLADHYVGAILLPTGA								
	FAYVILRAADQVCGDLIELSLITPLVPLATGAVOYIANGPDEARSRVYSCSD								
	DAVQDQSTCHAGTGLTSPDHODAGQDGPDIWPNDAVAVPDSFYAAREGFGDGR								
	AFQGLQAAKRGDEITFAEVLGILPTAHREDAGFGGIL								

[illegible]

ITAMGGYFRMTWTPTPIGAPLTLEVRTEDCVADEPGRPWTDSDMRPGWTLADATVQV
 PGDHSMDHEHAGSTAQAVASWLDKLNRTAROR"

BASE COUNT 2256 a 4805 c 6457 g 3249 t

ORIGIN

Query Match 40.0%; Score 40.4; DB 6; Length 16767;
 Best Local Similarity 63.3%; Pred. No. 18;
 Matches 62; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

OY 1 GTGCAGCCGAGATCGCGCGCTTTCATGGGGCCGCGAGCGCCGCGCGCGTGCAT 60
 ||| ||||| ||||| ||| ||| ||||| ||| |||||
 Db 15661 GTCCGCGCCGATATCCGGGTGACTGGGCGACGCGAGTCGAGCGATGTGACATCGAG 15720

OY 61 CGGCGCGCTGTGCGACTTGGGCTCGACTCGCTCAGCGC 98
 ||| ||| ||| ||| ||||| ||||| |||||
 Db 15721 AAGCCTTTGGCGAGCGTGGTTTCGACTCGCTCAGCGC 15758

RESULT 15
 AX089418

LOCUS AX089418 25360 bp DNA linear PAT 21-MAR-2001
 DEFINITION Sequence 3 from Patent WO0116303.
 ACCESSION AX089418
 VERSION AX089418.1 GI:13443679

KEYWORDS
 SOURCE Saccharopolyspora spinosa.
 ORGANISM Saccharopolyspora spinosa.

REFERENCE 1 (bases 1 to 25360)
 AUTHORS Eberz,G., Moehrle,V., Froede,R., Velten,R. and Salas,J.A.
 TITLE Nucleic acids which code for the enzyme activities of the spinosyn
 biosynthesis

JOURNAL Patent: WO 0116303-A 3 08-MAR-2001;
 BAYER AG (DE)

FEATURES
 source location/Qualifiers
 1..25360

BASE COUNT 3832 a 9143 c 8354 g 4031 t
 ORIGIN /organism="Saccharopolyspora spinosa"
 /db_xref="taxon:60894"

Query Match 40.0%; Score 40.4; DB 6; Length 25360;
 Best Local Similarity 63.3%; Pred. No. 17;
 Matches 62; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

OY 1 GTGCAGCCGAGATCGCGCGCTTTCATGGGGCCGCGAGCGCCGCGCGTGCAT 60
 ||| ||||| ||||| ||| ||| ||||| ||| |||||
 Db 92 GTCCGCGCCGATATCCGGGTGACTGGGCGACGCGAGTCGAGCGATGTGACATCGAG 151

OY 61 CGGCGCGCTGTGCGACTTGGGCTCGACTCGCTCAGCGC 98
 ||| ||| ||| ||| ||||| ||||| |||||
 Db 152 AAGCCTTTGGCGAGCGTGGTTTCGACTCGCTCAGCGC 189

Search completed: November 6, 2002, 13:17:56
 Job time : 618.625 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 6, 2002, 11:51:54 ; Search time 1197.62 Seconds
(without alignments)
1365.823 Million cell updates/sec

Title: US-09-724-876-2_COPY_30550_30650

Perfect score: 101

Sequence: 1 cggcggagcggccgtacag.....gcttcgcgggtgcgca 101

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: em_estba:*
2: em_esthum:*
3: em_estlin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hlc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_man:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rnd:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Match Length	ID	Description
1	34.2	33.9	786	9	AL563086
2	31.4	30.9	1158	13	BM465130
3	31.2	30.9	500	9	AA464685
4	30.8	30.5	910	17	CNS03587
5	30.4	30.1	673	12	BG810602
6	30.2	29.9	819	17	A0840535

C	7	30	29.7	423	14	W79688
8	30	29.7	427	13	Bj235954	
9	30	29.7	457	13	B1468728	
10	30	29.7	641	13	B1062639	
11	30	29.7	924	12	BG681148	
12	29.8	29.5	611	10	AW972109	
13	29.8	29.5	935	13	B1100110	
14	29.6	29.3	327	13	B1721001	
15	29.6	29.3	378	10	AW415295	
16	29.6	29.3	967	14	BA429819	
17	29.6	29.3	1059	12	BE730429	
18	29	28.7	534	17	TA293G09Q	
19	29	28.7	835	17	CNS02A4S	
20	29	28.7	988	13	B1246425	
21	29	28.7	1018	14	B0715891	
22	28.8	28.5	420	9	AJ475363	
23	28.8	28.5	610	10	AV941778	
24	28.8	28.5	861	13	B1818720	
25	28.6	28.3	437	9	A1816540	
26	28.6	28.3	508	17	AO867670	
27	28.6	28.3	514	17	P120L	
28	28.6	28.3	640	14	BM690391	
29	28.4	28.1	445	9	AA076633	
30	28.4	28.1	486	13	BM087592	
31	28.4	28.1	520	12	BF628130	
32	28.4	28.1	529	13	B1847926	
33	28.4	28.1	644	13	B1462568	
34	28.4	28.1	651	9	A1971742	
35	28.4	28.1	669	17	CNS045SK	
36	28.4	28.1	836	12	BF623428	
37	28.4	28.1	1008	17	CNS05CO2	
38	28.2	27.9	245	17	BH199283	
39	28.2	27.9	349	17	AO907893	
40	28.2	27.9	399	17	BH193944	
41	28.2	27.9	406	10	AM297140	
42	28.2	27.9	447	10	BE219371	
43	28.2	27.9	456	9	A1091672	
44	28.2	27.9	486	9	A1215893	
45	28.2	27.9	494	9	AI305222	

ALIGNMENTS

RESULT 1
LOCUS AL563086 786 bp mRNA linear EST 16-FEB-2001
DEFINITION AL563086 LTI_NFL003_NBC3 Homo sapiens CDNA clone CSODC025YL03 3
ACCESSION AL563086
VERSION AL563086.1 GI:12912151
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 786)
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE Full-length CDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES

source
1..786
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CSODC025YL03"
/clone_id="LTI_NFL003_NBC3"
/sex="male"
/tissue_type="neuroblastoma cells"
/lab_host="DH10B"

/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact: Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax: (1) 301 610 8371 Email: fliang@lifetech.com URL: <http://fulllength.invitrogen.com>"

BASE COUNT 158 a 221 c 216 g 177 t 14 others
ORIGIN

Query Match 33.9%; Score 34.2; DB 9; Length 786;
Best Local Similarity 63.0%; Pred. No. 12;
Matches 51; Conservative 1; Mismatches 29; Indels 0; Gaps 0;

QY 21 GCGGTATCCATCCAGCGGCGTCGATGCTGCTTGATGACGACCCAGC 80
DB 181 GGAGTTATMTTGGCATGTCACGACCTATGCGGCTGCTGTGAGCAGGACGCTA 240
QY 81 CGCTCTCGCGGGTCTGCGCA 101
DB 241 CTCCTCGCGCGTGTGCTGCA 261

RESULT 2
LOCUS BM465130 1158 bp mRNA linear EST 05-FEB-2002
DEFINITION AGENCOUNT_6427971 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:5501855
5', mRNA sequence.
ACCESSION BM465130
VERSION BM465130.1 GI:18514172
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 1158)
AUTHORS NIH-MGC <http://mhc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: c9apbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Plate: LAM12139 row: a column: 24
High quality sequence start: 115
High quality sequence stop: 671.
Location/Qualifiers

FEATURES
source 1..1158
Location/Qualifiers

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5501855"
/clone_lib="NIH_MGC_67"
/tissue_type="Retinoblastoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.75 kb. Library constructed by Life Technologies."
technology

BASE COUNT 243 a 364 c 328 g 190 t 33 others
ORIGIN

Query Match 31.1%; Score 31.4; DB 13; Length 1158;
Best Local Similarity 61.7%; Pred. No. 79;
Matches 50; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 21 GCGGTATCCATCCAGCGGCGTCGATGCTGCTTGATGACGACCCAGC 80
DB 733 GGAGTTATMTTGGCATGTCACGACCTTGGGCTGCTGTGAGCAGGACGCTA 674
QY 81 CGCTCTCGCGGGTCTGCGCA 101
DB 673 CTCCTCGCGCGTGTGCTGCA 653

RESULT 3
LOCUS AA464685 500 bp mRNA linear EST 10-JUN-1997
DEFINITION zx85f11.r1 Soares ovary tumor NbHOT Homo sapiens cDNA clone
IMAGE:810573 5', mRNA sequence.
ACCESSION AA464685
VERSION AA464685.1 GI:2189569
KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 500)
AUTHORS Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, B., Schellenberg, K., Stepien, M., Tan, F., Theisling, B., White, Y., Wylie, T., Waterston, R. and Wilson, R.
WashU-Merck EST Project 1997
JOURNAL Unpublished (1997)
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu

This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -28m3 rev2 ET from Amersham
High quality sequence stop: 458.
Location/Qualifiers

FEATURES
source 1..500
Location/Qualifiers

/organism="Homo sapiens"
/db_xref="GDB:6041102"
/db_xref="taxon:9606"
/clone="IMAGE:810573"
/clone_lib="Soares ovary tumor NbHOT"
/sex="Female"
/tissue_type="ovarian tumor"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: ovary; Vector: pT73D (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACATCGAAGTGGAGCGCGCGGTTTTTTTTTTT 3'] double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 84 a 188 c 133 g 94 t 1 others
ORIGIN

Query Match 30.9%; Score 31.2; DB 9; Length 500;
Best Local Similarity 66.2%; Pred. No. 73;
Matches 45; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 33 GCAGCGGAGCGCTCGATGATGCTGTGATGACGACCCAGCGCTTCGCGG 92
DB 339 GCAGCGGAGCGCTCGTGGCCAGTGGTGTGAGCTGCATGACGAGCGCTTGGGG 280

QY 93 GTGCTGC 100
DB 279 GTGTGGC 272

REFERENCE	1 (bases 1 to 673)
AUTHORS	Choi, W. and Dean, R. A.
TITLE	Construction and sequence analysis of an appressorium stage cDNA library in the rice blast fungus, <i>Magnaporthe grisea</i> Unpublished (2001)
JOURNAL	Contact: Ralph A. Dean Fungal Genomics Laboratory North Carolina State University Campus Box 7251, Raleigh, NC 27695, USA Tel: 919-513-0020 Fax: 919-513-0024 Email: ralph.dean@ncsu.edu
COMMENT	Seq primer: T3 primer (AATTAACTCCTACTAAGAAGG). Location/Qualifiers
FEATURES	1. 673 /organism="Magnaporthe grisea" /strain="70-15" /db_xref="taxon:148305" /clone="mgct006x013f" /clone_id="Magnaporthe grisea Appressorium Stage cDNA" /dev_stage="germinated conidia on appressorium-inductive surface" /note="Vector: pBluescript SK(+). Vector: Site 1: EcoRI; Site 2: XhoI. The appressorium formation-specific cDNA library was constructed from conidia germinated for 5-8 hr on an inductive surface. The library contains over 55,000 clones with average insert size of 1.5 kbp."
BASE COUNT	154 a 232 c 170 g 117 t
ORIGIN	
Query Match	30.1%; Score 30.4; DB 12; Length 673;
Best Local Similarity	57.3%; Pred. No. 1.3e+02;
Matches	55; Conservative 0; Mismatches 41; Indels 0; Gaps 0;
QY	2 GGGCGAGTGGCGGTATACAGGCGGTGATCCAGACGAGCGGTCGATGATGGTGGCT 61
Db	411 GGTGAAGTGGCGGCTAGAGGGTGTGTGTGCTGATGTGCTCGCGCTGATGTGTGTC 352
QY	62 TGATGACGACAGCACCGACCGCTTCTCGCGGGTCT 97
Db	351 TGAATGTGCTGCTGCCTGACGTCGTGCTGCTGTGTGT 316
RESULT 6	
LOCUS	A0840535 819 bp DNA linear GSS 01-SEP-1999
DEFINITION	nxhdb0064804r CUGI Rice BAC Library Oryza sativa genomic clone
ACCESSION	nxhdb0064804r, DNA sequence.
VERSION	A0840535
KEYWORDS	A0840535.1 GI:5818588
SOURCE	GSS.
ORGANISM	Oryza sativa. Oryza sativa. Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE	1 (bases 1 to 819)
AUTHORS	Wing, R. A. and Dean, R. A.
TITLE	A BAC End Sequencing Framework to Sequence the Rice Genome
JOURNAL	Unpublished (1998)
COMMENT	Contact: Wing RA Clemson University Genomics Institute Clemson University 100 Jordan Hall, Clemson, SC 29634, USA Tel: 864 656 7288 Fax: 864 656 4293 Email: rwing@clemson.edu Seq primer: GGAAACAGCTATGACCATG Class: BAC ends High quality sequence stop: 194. Location/Qualifiers
FEATURES	1. 819 /organism="Oryza sativa"
SOURCE	

```

/strain="japonica"
/cultivar="Nipponbare"
/db_xref="taxon:4530"
/clone="nbp0064B04r"
/clone_lib="CUGI Rice BAC library"
/tissue_type="leaf"
/lab_host="E. coli DH10B"
/notes="Vector: pBelBAC11; Site_1: HindIII; Site_2: HindIII; Rice is one of two most popular grains in the world. Half of the world population especially those inhabiting highly populated areas of the humid tropics and subtropics, rely on rice as their primary source of carbohydrate. Monocotyledonous rice is a diploid plant (2n=24) with a haploid genome equivalent of 431 Mbp (Arumuganathan and Earle, 1991). The relatively small genome of rice, three times larger than that of Arabidopsis, makes it suitable for genomic studies. In order to facilitate positional cloning, physical mapping and genome sequencing of rice, we have constructed a BAC library from Oryza sativa, Nipponbare variety. The library contains 36,864 clones with an average insert size of 128.5 Kb providing 10.9 haploid genome equivalents. The deep coverage allows the isolation a particular sequence with a probability of 99.9 %. Two high density filters, each containing 18,432 clones (doubly spotted), represent the whole library for colony screening."

BASE COUNT      174 a      216 c      198 g      231 t
ORIGIN

```

```

Query Match      29.9%; Score 30.2; DB 17; Length 819;
Best Local Similarity 69.5%; Pred. No. 1.6e+02;
Matches 41; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

```

```

QY      5  GGAGTGGCGCTTACAGGCGGTATCATGCAGCGCGCTGATGATGATGCTGCTG 63
      1111 111111 111111 111 111111 1111 1111 111111
Db      297 GGAGCTGCGCTTCCAGGCGCTCTCCCTCGACGACGACGATGCTCTCTG 355

```

```

RESULT 7
LOCUS      W79688      423 bp      mRNA      linear      EST 17-OCT-1996
DEFINITION zdbec01.r1 Soares fetal heart NbHH19W Homo sapiens cDNA clone
ACCESSION W79688
VERSION W79688.1 GI:1390097
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Chordata; Granulata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman
M., Hultman, M., Kucuba, T., Le, M., Lennon, G., Marra, M., Parsons, J.,
Rifkin, B., Rohlfing, T., Soares, M., Tan, F., Trevaskis, E., Waterston
R., Williamson, A., Wohldmann, P. and Wilson, R.
The WashU-Merck EST Project
Unpublished (1995)
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert Length: 968 Std Error: 0.00
Seq primer: mob.RGA+ET
High quality sequence scop: 384.
Location/Qualifiers
1. 423

```

```

FEATURES
source
/organism="Homo sapiens"
/db_xref="GDB:1272895"
/db_xref="taxon:9606"

```

```

/clone="IMAGE:347520"
/clone_lib="Soares_fetal_heart_NbHH19W"
/sex="unknown"
/dev_stage="19 weeks"
/lab_host="DH10B (ampicillin resistant)"
/notes="Organ: heart; Vector: pT73D (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACATCTGATGAGTGGAGCGCGCATCTCTTTTCTTTTCTT 3']. TGTTCACATCTGATGAGTGGAGCGCGCATCTCTTTTCTTTTCTT 3']. double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by M. Fatima Bonaldo. This library was constructed from the same fetus as the fetal lung library, Soares fetal lung NBHH19W."

BASE COUNT      87 a      126 c      111 g      95 t
ORIGIN

```

```

Query Match      29.7%; Score 30; DB 14; Length 423;
Best Local Similarity 57.4%; Pred. No. 1.5e+02;
Matches 54; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

```

```

QY      8  GTGGCGCTTACAGGCGGTATCCATCCAGCGGCGCTGATGATGCTGCTGATGA 67
      1111 111111 1111 111 111111 1111 111 111 111
Db      250 GTGGGCGCTTACAGGTGGAGAGATGAGAGCGCACATGCGAGGAGGCAAGA 191
QY      68  GCAGACCACCGACCGCTTCGCGGGGTCTGGCA 101
      111111 111111 111111 111111 111111
Db      190 GCAGACTCACTGCTAGTGTACACAGTGGCTGCA 157

```

```

RESULT 8
LOCUS      BJ235954      427 bp      mRNA      linear      EST 05-APR-2002
DEFINITION BJ235954 Y. Ogihara unpublished cDNA library, whe Triticum aestivum cDNA clone whe19c05 5', mRNA sequence.
ACCESSION BJ235954
VERSION BJ235954.1 GI:20052145
KEYWORDS EST.
SOURCE bread wheat.
ORGANISM Triticum aestivum
REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
AUTHORS Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae
Ogihara, Y. and Murai, K.
Expressed genes in Triticum aestivum
Unpublished (2002)
COMMENT Contact: Tadao Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshin@genetics.nig.ac.jp.

```

```

FEATURES
source
Location/Qualifiers
1. 427

```

```

/organism="Triticum aestivum"
/cultivar="Chinese Spring"
/db_xref="taxon:4565"
/clone="whe19c05"
/clone_lib="Y. Ogihara unpublished cDNA library, whe"
/tissue_type="seed DRA10"
/dev_stage="seed DRA10"
/notes="Vector: Lambda Uni-ZAP XR, excised phagemid; Site_1: EcoRI; Site_2: XhoI; Plants were grown under hydroponic conditions at UC Davis, salt stressed for 12 hours, and for 7 days, then dissected and frozen (Akhunov in J. Dvorak Lab). Total RNA was prepared from sheath tissue, equal quantities of RNA were pooled from the two samples, polyA was purified from the pooled RNA, a cDNA

```


BASE COUNT	ORIGIN	131 a	195 c	189 g	125 t	1 others
tissue mRNA and cDNA amplification were performed under low stringency conditions."						

Query Match	29.7%	Score 30:	DB 13:	Length 641:
Best Local Similarity	57.4%	Pred. No. 1.7e+02:		
Matches 54:	Conservative 0:	Mismatches 40:	Indels 0:	Gaps 0

Qy 5 GGAGTGGCCGTTACAGSGCGTGATTCATGCAGCCGGAGGCCGTGATGATGCTGTGTA 64
+ + + + + + + + + + + + + + + + + +
Db 433 GAAGTGACGGGTAAGGGGGTGACAATTGAGCAGCGGGAGCTGGATGAATAATGCCCTGC 49

0y 65 TGAGCAGACCACCGGCTTCTCGGGGTGTG 98
 || | || || || | | | | |
 Db 493 AAGAATCCCTAACGCTGTCTTTGGCCGGCAGT 526

| | |
|-----------------|--|
| RESULT 11 | |
| RG681148/c | |
| LOCUS | |
| RG681148 | |
| 924 bp | |
| mpna | |
| 14near | |
| EST_01-MAY-2001 | |

| | | | | | |
|------------|------------------------------|--------------|------------|---------------|---|
| DEFINITION | 0020829/82.1 NC_01_CGAP_Skn4 | homo sapiens | cDNA clone | IMAGE:4753598 | 5 |
| ORIGIN | mRNA sequence. | | | | |
| ACCESSION | BC681148 | | | | |
| VERSION | BC681148.1 | GI:13912545 | | | |

| ALPHANUMERIC | SOURCE | ORGANISM |
|--------------|---|----------|
| ESI. | human. | |
| | Homo sapiens | |
| | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | |

REFERENCE
AUTHORS
TITLE

1 (bases 1 to 924)
NIH-MGC <http://mgc.ncl.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)

COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: James Cleaver, M.D.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be

http://image.llnl.gov
Plate: LLAM10613 row: 1 column: 19
High quality sequence stop: 839.

```
source      1. .924  
            /organism="Homo sapiens"  
            /db_xref="taxon:9606"
```

```

/clone_idb="NCI_CGAP_Skn4"
/tissue_type="squamous cell carcinoma"
/lab_host="DH10B (T1 phage-resistant)"
host="Dorana-4133 Vector"
chr="chr6"
site="1"

```

BASE COUNT
185 A
261 C
255 G
227 T

| ORIGIN | Query Match | Score | DB | Length |
|-----------------------|-------------|-------------|-------|--------|
| Best Local Similarity | 57.4% | 30 | 12 | 924 |
| | | Pred. NO. 1 | 1e+02 | |

0y 8 GTGGCCGTTACAGGCGTGATCCATGACGCGGAGCGCTCGATGATGCTGCTTGAATGA 67
 matches 24; conservative 0; mismatches 40; indels 0; gaps

DZ 370 GAGGGGACCTTCCATTTTCATTGCGGTCTGCA 101
| | | | |
DY 68 GCAGACCACCGACCGCTTCTGCGGTGCTGCCA 101
| | | | |

[illegible]

| | |
|------------|--|
| RESULT 12 | |
| AW972109 | |
| LOCUS | 611 bp mRNA linear |
| DEFINITION | EST384093 MACE resequences, MAGL Homo sapiens cDNA, mRNA sequence. |

```

VERSION      AM972109.1  GI:8161850
KEYWORDS
SOURCE       human.

```

| REFERENCE | Phases | Species | Number of individuals |
|---|-----------------|---------|-----------------------|
| Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Euthalia; Primates; Catarrhini, Hominoidea; Homo. 1 (phases 1 to 611) | Phases 1 to 611 | Homo | 1 |

| TITLE | Author |
|---|---------------------------------------|
| Assessment of gene expression patterns in a model of colon tumor metastatic seeding | 'I.E., Saeed,A.I., Quackenbush,J. |
| 10,200 element cDNA microarray | Sharov,V., Lee,N.H., Yeatman,T.J. and |

JOURNAL COMMENT
Unpublished (2000)
Contact: John Quackenbush
The Institute for Genomic Research
9713 Medical Center Dr. Rockville
MD 20850 USA

Tel: 301 838 3528
Fax: 301 838 0208
Email: johnq@tigr.org
Plate: 299

```

seq primer: forward.
FEATURES
    source
        location/Qualifiers
        1..611
        /organism="Homo sapiens"

```

| | |
|------------|-------|
| BASE COUNT | 164 a |
| | 143 c |
| | 173 q |
| | 131 r |

| | | | | |
|-----------------------|-------|-------------------|-------|------------|
| Query Match | 29.5% | Score 29.8 | DB 10 | Length 611 |
| Best Local Similarity | 62.4% | Pred. No. 1.9e+02 | | |

QY 1 CGGGGAGTGGCCCGTTACAGGGCGGTATTCATTGCAGCCGGAGCGCTGCATC-ATGGTG TG 59

60 CTTGATGAGCAGACCACCGCTTCTCGCGGCTGTCGC 100
||||| ||||| | | | | | | | | | | | | | | | |

RESULT 13

| | | | | | |
|-------------|--|--------|------|--------|-----------------|
| LOCUS | B1100110 | 935 bp | mRNA | linear | EST 26-JUN-2001 |
| DEFINITION | 602885179p1 NCL CGAP_K1d14 Mus musculus CDNA clone IMAGE:5040595 | | | | |
| DESCRIPTION | 5' mRNA sequence. | | | | |
| ACCESSION | | | | | |

ORGANISM house mouse.
Mus. musculus

REFERENCE
AUTHORS
NIH-MGC brtn://mgc.nci.nih.gov/
1 (bases 1 to 935)
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (bases 1 to 935)

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: ccabbs-remail.nih.gov

Issue Procurement: Jeffrey E. Green, M.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.

CLONE ASSISTANCE: Mac Clone assistance information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>

polya minn was purified from each sample, pooled and cDNA synthesized. The cDNA was directionally cloned into lambda

CCGTTACAGGGCGTGATCCATGCAGCCGGAGCGCTCGATGATGGTGCT 61

polya minn was purified from each sample, pooled and cDNA synthesized. The cDNA was directionally cloned into lambda

Db 85 GGAGATTGGCCAGAGAGGTCACAAACAGAGAGTCGAGAGTTGGGTGGTGCCCT 144
QY 62 TGATGAGCAGACCACC 77
||| |||
Db 145 GAGCACACAGAGCACC 160

Search completed: November 6, 2002, 15:52:22
Job time : 1202.62 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 6, 2002, 13:32:49 : Search time 27.125 Seconds
(without alignments)
1240.503 Million cell updates/sec

Title: US-09-724-876-2_COPY_30550_30650

Perfect score: 101

Sequence: 1 cgcgcgagtgccgtacag.....gtcttcgcggtcgtgca 101

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 310279 seqs, 16577418 residues

Total number of hits satisfying chosen parameters: 620558

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications_NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/2/pubpna/PCr_NEM_PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEM_PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEM_PUB.seq:*
- 6: /cgn2_6/ptodata/2/pubpna/PCrUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEM_PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubpna/US09_NEM_PUB.seq:*
- 10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/2/pubpna/US10_NEM_PUB.seq:*
- 12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/2/pubpna/US60_NEM_PUB.seq:*
- 14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length DB | ID | Description |
|------------|-------|-------------|-----------|----|---------------------|
| 1 | 37 | 36.6 | 13842 | 10 | US-09-861-289-30 |
| 2 | 37 | 36.6 | 36778 | 10 | US-09-861-289-5 |
| 3 | 26.8 | 26.5 | 263 | 10 | US-09-923-876-3525 |
| 4 | 26.6 | 26.3 | 202 | 10 | US-09-815-242-1105 |
| 5 | 26.6 | 26.3 | 443 | 10 | US-09-815-242-1258 |
| 6 | 26.6 | 26.3 | 540 | 10 | US-09-815-242-4112 |
| 7 | 26.4 | 26.1 | 538 | 10 | US-09-771-161A-70 |
| 8 | 26.4 | 26.1 | 2457 | 12 | US-10-052-586-569 |
| 9 | 26.4 | 26.1 | 13224 | 10 | US-09-764-853-897 |
| 10 | 26 | 25.7 | 260 | 10 | US-09-815-242-1239 |
| 11 | 26 | 25.7 | 536 | 10 | US-09-873-880-21 |
| 12 | 26 | 25.7 | 1254 | 10 | US-09-815-242-6157 |
| 13 | 26 | 25.7 | 15872 | 10 | US-09-861-289-1 |
| 14 | 26 | 25.7 | 32204 | 10 | US-09-764-855-327 |
| 15 | 25.8 | 25.5 | 2968 | 10 | US-09-815-915-1 |
| 16 | 25.6 | 25.3 | 412 | 10 | US-09-860-352-12815 |
| 17 | 25.6 | 25.3 | 414 | 10 | US-09-960-352-12671 |
| 18 | 25.6 | 25.3 | 429 | 10 | US-09-960-352-7004 |
| 19 | 25.6 | 25.3 | 1501 | 10 | US-09-768-894A-1 |

| | | | | | | |
|----|------|------|-------|----|---------------------|-------------------|
| 20 | 25.6 | 25.3 | 6503 | 10 | US-09-935-541-12 | Sequence 12, Appl |
| 21 | 25.4 | 25.1 | 395 | 10 | US-09-867-701-6689 | Sequence 6689, Ap |
| 22 | 25.4 | 25.1 | 419 | 10 | US-09-867-701-6858 | Sequence 6858, Ap |
| 23 | 25.4 | 25.1 | 1561 | 10 | US-09-729-674-177 | Sequence 177, App |
| 24 | 25.4 | 25.1 | 1677 | 10 | US-09-915-582-16 | Sequence 16, Appl |
| 25 | 25.4 | 25.1 | 1802 | 10 | US-09-799-777-104 | Sequence 104, App |
| 26 | 25.4 | 25.1 | 3012 | 10 | US-09-767-215-3 | Sequence 3, Appl1 |
| 27 | 25.4 | 25.1 | 3409 | 10 | US-09-954-456-501 | Sequence 501, App |
| 28 | 25.4 | 25.1 | 3417 | 10 | US-09-767-215-4 | Sequence 4, Appl1 |
| 29 | 25.4 | 25.1 | 3931 | 10 | US-09-767-215-1 | Sequence 1, Appl1 |
| 30 | 25.2 | 25.0 | 232 | 10 | US-09-864-761-31781 | Sequence 31781, A |
| 31 | 25.2 | 25.0 | 512 | 10 | US-09-864-761-15259 | Sequence 15259, A |
| 32 | 25.2 | 25.0 | 1738 | 10 | US-09-764-877-3218 | Sequence 3218, Ap |
| 33 | 25 | 24.8 | 1848 | 10 | US-09-915-582-34 | Sequence 34, Appl |
| 34 | 24.8 | 24.6 | 398 | 10 | US-09-960-352-1749 | Sequence 1749, Ap |
| 35 | 24.6 | 24.4 | 531 | 10 | US-09-864-761-6877 | Sequence 6877, Ap |
| 36 | 24.6 | 24.4 | 1845 | 10 | US-09-822-845A-292 | Sequence 292, App |
| 37 | 24.4 | 24.2 | 263 | 10 | US-09-878-574-15767 | Sequence 15767, A |
| 38 | 24.4 | 24.2 | 933 | 9 | US-09-934-392-3 | Sequence 3, Appl1 |
| 39 | 24.4 | 24.2 | 1435 | 10 | US-09-813-319A-1 | Sequence 1, Appl1 |
| 40 | 24.4 | 24.2 | 1543 | 9 | US-09-934-392-1 | Sequence 1, Appl1 |
| 41 | 24.4 | 24.2 | 1566 | 10 | US-09-833-381-987 | Sequence 987, App |
| 42 | 24.4 | 24.2 | 1700 | 10 | US-09-817-184-1 | Sequence 1, Appl1 |
| 43 | 24.4 | 24.2 | 3209 | 10 | US-09-925-301-474 | Sequence 474, App |
| 44 | 24.4 | 24.2 | 24881 | 10 | US-09-817-184-3 | Sequence 3, Appl1 |
| 45 | 24.4 | 24.2 | 29695 | 10 | US-09-752-820A-3 | Sequence 3, Appl1 |

ALIGNMENTS

RESULT 1

US-09-861-289-30

Sequence 30, Application US/09861289

Patent No. US20020110897A1

GENERAL INFORMATION:

APPLICANT: Sherman, D.H.

APPLICANT: Liu, H.

APPLICANT: Xue, Y.

APPLICANT: Zhao, L.

TITLE OF INVENTION: DNA encoding methymycin and pikromycin

FILE REFERENCE: 600.438051

CURRENT APPLICATION NUMBER: US/09/861,289

PRIOR APPLICATION NUMBER: 2001-05-18

PRIOR FILING DATE: 1998-06-26

NUMBER OF SEQ ID NOS: 43

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 30

LENGTH: 13842

TYPE: DNA

ORGANISM: Streptomyces venezuelae

US-09-861-289-30

Query Match 36.6%; Score 37; DB 10; Length 13842;

Best Local Similarity 62.4%; Pred. No. 0.0046;

Matches 58; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

QY 1 CGCGGAGTGGCCGTTACAGGCGGTATCATGACGCGAGCGCTGATGATGCTGTC 60

DB 6848 CCGCGGAGACGCGCCCTTCACCGCGCTGTCACACCGCGCGCGCTGACGACGCGCATCG 6907

QY 61 TTGATGACGACGACCGACCGACCGCGCTTCTGCGGG 93

DB 6908 TGGACACGCTGACCGCGACGACGATCGCGGG 6940

RESULT 2

US-09-861-289-5

Sequence 5, Application US/09861289

Patent No. US20020110897A1

GENERAL INFORMATION:

APPLICANT: Sherman, D.H.

```

; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600 438US1
; CURRENT APPLICATION NUMBER: US/09/861,289
; PRIOR FILING DATE: 2001-05-18
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO: 5
; LENGTH: 36778
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-861-289-5
```

```

Query Match          36.6%; Score 37; DB 10; Length 36778;
Best Local Similarity 62.4%; Pred. No. 0.0054;
Matches 58; Conservative 0; Mismatches 35; Indels 0; Gaps 0;
```

```

QY 1 CGCGGAGTGGCGTTACAGAGGCGGTATCCATGACGCCGGAGCGTCGATGATGGTGTGC 60
DB 8589 CGCGCGAGAGCGCCCTCACCCTCGCTCCACACCGCGCGCTCGACGAGGCAATCG 8648
QY 61 TTGATGACGACGACCGACCGCGCTTCGCGCG 93
DB 8649 TGGACACGCTGACCGCGGACGACGAGTCCGCGCGG 8681
```

```

RESULT 3
US-09-923-876-3525/c
; Sequence 3525, Application US/09923876
; Patent No. US20020013958A1
; GENERAL INFORMATION:
; APPLICANT: Laligudi, Raghunath V.
; APPLICANT: Kamigaki, Laura Y. (Ito)
; APPLICANT: Sherman, Bradley K.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN SEEDLING
; FILE REFERENCE: PL-0012-1 CON
; CURRENT APPLICATION NUMBER: US/09/923,876
; CURRENT FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: 09/298,329
; PRIOR FILING DATE: 1999-04-21
; PRIOR APPLICATION NUMBER: 60/085,331
; PRIOR FILING DATE: 1998-05-05
; NUMBER OF SEQ ID NOS: 6332
; SOFTWARE: PERL Program
; SEQ ID NO: 3525
; LENGTH: 263
; TYPE: DNA
; ORGANISM: Zea mays
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20020013958A1 700162447H1
US-09-923-876-3525
```

```

Query Match          26.5%; Score 26.8; DB 10; Length 263;
Best Local Similarity 55.3%; Pred. No. 4.5;
Matches 52; Conservative 0; Mismatches 42; Indels 0; Gaps 0;
```

```

QY 5 GGAGTGGCGGTTACAGAGGCGGTATCCATGACCGGAGCGCTCGATGATGGTGTGA 64
DB 95 GCGGAGCGGTGTCAGTAGAGAGCGAGCTAAGAAAGGCGGAGGCGGAGGCTTTCGCG 36
QY 65 TGAGCAGACCGACCGACCGCTTCGCGGGGTCTG 98
DB 35 TGGCGGCTTCGACCGCAGATTGGGCGGCTG 2
```

```

RESULT 4
US-09-815-242-1105
; Sequence 1105, Application US/09815242
```

```

; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 1105
; LENGTH: 202
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-815-242-1105
```

```

Query Match          26.3%; Score 26.6; DB 10; Length 202;
Best Local Similarity 58.0%; Pred. No. 5;
Matches 47; Conservative 0; Mismatches 34; Indels 0; Gaps 0;
```

```

QY 6 GAGTGGCGGTTACAGAGGCGGTATCCATGACCGGAGCGCTCGATGATGGTGTGAT 65
DB 68 GAAAGCAGCGAGCGAGCGGCTTCGTCATCCGTACGAGCGGTGTGTCGATGAT 127
QY 66 GAGCAGACCGACCGACCGCTTC 86
DB 128 GTCCAGGCCACGACGCGCATC 148
```

```

RESULT 5
US-09-815-242-1258
; Sequence 1258, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
```

```
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1258
; LENGTH: 443
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-815-242-1258

Query Match      26.3%; Score 26.6; DB 10; Length 443;
Best Local Similarity 58.0%; Pred. No. 5.6;
Matches 47; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

OY      6 GAGTGGCCGTTACAGGCGGTGATCCAGCCGAGCGCTCGATGATGTTGCTTGT 65
        || | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      76 GAAAGCAGCGACGAGGCGCGACCTTCGTATCCGTACGACGCGTGTGTACGGGTGAT 135

OY      66 GAGCAGACCACCGACCGCTTC 86
        | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      136 GTCCAGGCCACGACGCGCATC 156

RESULT 6
US-09-815-242-4112/c
; Sequence 4112, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4112
; LENGTH: 540
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-815-242-4112

Query Match      26.3%; Score 26.6; DB 10; Length 540;
Best Local Similarity 58.0%; Pred. No. 5.7;
Matches 47; Conservative 0; Mismatches 34; Indels 0; Gaps 0;
```

```
OY      6 GAGTGGCCGTTACAGGCGGTGATCCAGCCGAGCGCTCGATGATGTTGCTTGT 65
        || | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      519 GAAAGCAGCGACGAGGCGCGACCTTCGTATCCGTACGACGCGTGTGTACGGGTGAT 460

OY      66 GAGCAGACCACCGACCGCTTC 86
        | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      459 GTCCAGGCCACGACGCGCATC 439

RESULT 7
US-09-771-161A-70/c
; Sequence 70, Application US/09771161A
; Patent No. US20020110811A1
; GENERAL INFORMATION:
; APPLICANT: LEVINE, et al.
; TITLE OF INVENTION: VARIANTS OF PROTEIN KINASES
; FILE REFERENCE: 802620-2005.1
; CURRENT APPLICATION NUMBER: US/09/771,161A
; PRIOR FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: 09/724,676
; PRIOR FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: 136776
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: 135619
; PRIOR FILING DATE: 2000-04-12
; NUMBER OF SEQ ID NOS: 273
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 70
; LENGTH: 538
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: -
; LOCATION: (1)..(538)
; OTHER INFORMATION: "n" can be any nucleotide 'a', 'c', 'g' or 't'
US-09-771-161A-70

Query Match      26.1%; Score 26.4; DB 10; Length 538;
Best Local Similarity 65.0%; Pred. No. 6.6;
Matches 39; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

OY      38 CGAGCGCTCGATGATGTTGCTTGTGATGAGCAGCAGCCGCTTCGCGGCTCT 97
        || | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      377 CGAGCGAAGATTTGGGTGATTTGAAGAGCGCATGCTGCTTCGCTGCGCTT 318

RESULT 8
US-10-052-586-569
; Sequence 569, Application US/10052586
; Patent No. US20020127564A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C1
; CURRENT APPLICATION NUMBER: US/10/052,586
; PRIOR FILING DATE: 2002-01-15
; PRIOR APPLICATION NUMBER: 60/059263
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/059266
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/062250
```

[illegible]

```
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088811
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088824
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088825
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088826
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088861
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/088863
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/088876
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/089090
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: 60/089105
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: 60/089512
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089514
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089538
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089598
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089653
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089908

Query Match          26.1%; Score 26.4; DB 12; Length 2457;
Best Local Similarity 57.1%; Pred. No. 8.3;
Matches 48; Conservative 0; Mismatches 36; Indels 0; Gaps 0;
```

```
QY 5 GGATGGCGCTTACAGGCGTGATCCATGCAGCGCGCTCGATGATGTCCTTGA 64
Db 2126 GGAGGGGCGCTCCCGCGCTGCTGCGGAGCAGCAGTTTCTATGTGGAGCAGC 2185
```

```
QY 65 TGAGCAGACCCAGCCGCTTCTC 88
Db 2186 TGAGGAGGCGAGAACCCGCTCTC 2209
```

```
RESULT 9
US-09-764-853-897/c
; Sequence 897, Application US/09/64853
; Patent No. US20020090672A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PYZ06
; CURRENT APPLICATION NUMBER: US/09/764,853
; PRIOR FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 939
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 897
; LENGTH: 13224
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-853-897
```

```
Query Match          26.1%; Score 26.4; DB 10; Length 13224;
Best Local Similarity 69.2%; Pred. No. 11;
Matches 36; Conservative 0; Mismatches 16; Indels 0; Gaps 0;
```

```
QY 50 TGATGCTGTCTGATGAGCAGACACCGCTTCTCGGGGGGCTGCGCA 101
Db 11216 TGGGGTGTGCTGTGAGCAGGCGAGCTCTCTCGCGCGTGTGCTGCA 11165
```

RESULT 10

```
US-09-815-242-1259
; Sequence 1259, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: EUTRA.011a
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1259
; LENGTH: 260
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-815-242-1259
```

```
Query Match          25.7%; Score 26; DB 10; Length 260;
Best Local Similarity 59.5%; Pred. No. 8;
Matches 44; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

QY 13 CGTTACAGGGCGTGATCCATGCAGCGCGCTCGATGATGTCCTGATGAGCA 72
Db 2 CGCAGCAGGGCGCGACCTTCATCGTACGAGCGGTGTGTCAGGTCATGTCCAGG 61

QY 73 CCACCGACCGCTTC 86
Db 62 CCACCGACCGCATC 75
```

```
RESULT 11
US-09-873-880-21
; Sequence 21, Application US/09873880
; Patent No. US20020123118A1
; GENERAL INFORMATION:
; APPLICANT: Sewalt, Vincent
; APPLICANT: Falco, S. Carl
; APPLICANT: Allen, Stephen M.
; TITLE OF INVENTION: GLYCINE METABOLISM ENZYMES
; FILE REFERENCE: B81192 US CIP
; CURRENT APPLICATION NUMBER: US/09/873,880
; PRIOR FILING DATE: 2001-06-04
; PRIOR APPLICATION NUMBER: 09/363,321
; PRIOR FILING DATE: July 28, 1999
; PRIOR APPLICATION NUMBER: 60/094,839
; PRIOR FILING DATE: July 31, 1998
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 21
; LENGTH: 556
; TYPE: DNA
```

```

; ORGANISM: Trilicium aestivum
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (208)
; OTHER INFORMATION: n=A, C, G, or T
; NAME/KEY: unsure
; LOCATION: (330)
; OTHER INFORMATION: n=A, C, G, or T
; NAME/KEY: unsure
; LOCATION: (357)
; OTHER INFORMATION: n=A, C, G, or T
; NAME/KEY: unsure
; LOCATION: (360)
; OTHER INFORMATION: n=A, C, G, or T
; NAME/KEY: unsure
; LOCATION: (362)
; OTHER INFORMATION: n=A, C, G, or T
; NAME/KEY: unsure
; LOCATION: (389)
; OTHER INFORMATION: n=A, C, G, or T
; NAME/KEY: unsure
; LOCATION: (430)
; OTHER INFORMATION: n=A, C, G, or T
; NAME/KEY: unsure
; LOCATION: (437)
; OTHER INFORMATION: n=A, C, G, or T
; NAME/KEY: unsure
; LOCATION: (467)
; OTHER INFORMATION: n=A, C, G, or T
; NAME/KEY: unsure
; LOCATION: (491)
; OTHER INFORMATION: n=A, C, G, or T
; NAME/KEY: unsure
; LOCATION: (502)
; OTHER INFORMATION: n=A, C, G, or T
; NAME/KEY: unsure
; LOCATION: (515)
; OTHER INFORMATION: n=A, C, G, or T
; NAME/KEY: unsure
; LOCATION: (520)
; OTHER INFORMATION: n=A, C, G, or T
; NAME/KEY: unsure
; LOCATION: (548)
; OTHER INFORMATION: n=A, C, G, or T
; NAME/KEY: unsure
; LOCATION: (551)
; OTHER INFORMATION: n=A, C, G, or T
US-09-873-880-21

```

```

Query Match          25.7%; Score 26; DB 10; Length 556;
Best Local Similarity 53.5%; Pred. No. 9;
Matches 53; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

```

```

; RESULT 12
; US-09-815-242-6157/C
; Sequence 6157, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.

```

```

; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6157
; LENGTH: 1254
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1254)
US-09-815-242-6157

```

```

Query Match          25.7%; Score 26; DB 10; Length 1254;
Best Local Similarity 59.5%; Pred. No. 10;
Matches 44; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

```

```

; RESULT 13
; US-09-861-289-1
; Sequence 1, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438U51
; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 15872
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-861-289-1

```

```

Query Match          25.7%; Score 26; DB 10; Length 15872;
Best Local Similarity 54.1%; Pred. No. 15;
Matches 53; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

```

Db 7566 CCGCGAGCATCGCTGACGCGCCGTCGTCACACGCGGAGGTGTCAGCAGCGGAGCG 7625
QY 61 TTGATGAGACAGACACCGACCGCTTCTGCGGGTCTG 98
Db 7626 TGGAGGCGCTCACACCGGAGCGGCTGAGCGGCTACTG 7663

RESULT 14

US-09-764-855-327/c
; Sequence 327, Application US/09764855
; Patent No. US20020119919A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: P110
; CURRENT APPLICATION NUMBER: US/09/764,855
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 334
; SOFTWARE: Patent Ver. 2.0
; SEQ ID NO 327
; LENGTH: 32204
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-855-327

Query Match 25.7%; Score 26; DB 10; Length 32204;
Best Local Similarity 55.6%; Pred. No. 16;
Matches 50; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 3 GCGAGTGGCCGTTACAGGGGCGTGATCCATGACGCGGAGCGCTCGATGATGTGCTT 62
Db 102 GCGAGGCGCGTGTGCGGGTGCGTGGGTGCAAGGCGCGCGGAGGCGCGGGG 43
QY 63 GATGAGCAGACCGACCGACCGCTTCTCGCGG 92
Db 42 GCCGCGGAGGCGCCGAGACCGCTTCTGCGCG 13

RESULT 15

US-09-815-915-1/c
; Sequence 1, Application US/09815915
; Patent No. US20020025931A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; APPLICANT: Meyers, Rachel
; TITLE OF INVENTION: 3714, 16742, 23546, AND 13887 NOVEL
; TITLE OF INVENTION: PROTEIN KINASE MOLECULES AND USES THEREFOR
; FILE REFERENCE: 38155-20006.00
; CURRENT APPLICATION NUMBER: US/09/815,915
; CURRENT FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: US 60/191,846
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 2968
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (140)...(2491)
; NAME/KEY: misc_feature
; LOCATION: (1)...(2968)
; OTHER INFORMATION: n = A,T,C or G
US-09-815-915-1

Query Match 25.5%; Score 25.8; DB 10; Length 2968;
Best Local Similarity 56.5%; Pred. No. 13;
Matches 48; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 4 CGAGTGGCCGTTACAGGGGCGTGATCCATGACGCGGAGCGCTCGATGATGTGCTTG 63

Db 207 CGAGGGGCTTCTGTGCGCCCTGACCTGCGCGGGGTCGCGCTGAACCTCCGACATG 148
QY 64 ATGACGAGACCAACCGACCGCTTCTG 88
Db 147 ATAACCATGCTCCGCGGCGACCTC 123

Search completed: November 6, 2002, 20:30:50
Job time : 51.125 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 6, 2002, 11:51:54 : Search time 33.5 Seconds
(without alignments)
924.608 Million cell updates/sec

Title: US-09-724-876-2_COPY_30550_30650

Perfect score: 101

Sequence: 1 cggcgagtgccgttaccag.....gtcttcgagtgctgca 101

Scoring table: IDENTITY_NUC

Gap 10.0 , Gape 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Issued_Patents.NA:*

- 1: /cgn2_6/ptodata/2/ina/5A.COMB.seq:*
- 2: /cgn2_6/ptodata/2/ina/5B.COMB.seq:*
- 3: /cgn2_6/ptodata/2/ina/6A.COMB.seq:*
- 4: /cgn2_6/ptodata/2/ina/6B.COMB.seq:*
- 5: /cgn2_6/ptodata/2/ina/PCTUS.COMB.seq:*
- 6: /cgn2_6/ptodata/2/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|---------------------|--------------------|
| 1 | 101 | 100.0 | 68750 | 3 US-09-335-409-1 | Sequence 1, Appli |
| 2 | 101 | 100.0 | 68750 | 4 US-09-568-102-1 | Sequence 1, Appli |
| 3 | 101 | 100.0 | 68750 | 4 US-09-567-969-1 | Sequence 1, Appli |
| 4 | 101 | 100.0 | 68750 | 4 US-09-568-480-1 | Sequence 1, Appli |
| 5 | 101 | 100.0 | 68750 | 4 US-09-568-486-1 | Sequence 1, Appli |
| 6 | 101 | 100.0 | 68750 | 4 US-09-568-472-1 | Sequence 1, Appli |
| 7 | 101 | 100.0 | 68750 | 4 US-09-567-899-1 | Sequence 1, Appli |
| 8 | 101 | 100.0 | 71989 | 4 US-09-443-501A-2 | Sequence 1, Appli |
| 9 | 51.6 | 51.1 | 33529 | 4 US-09-144-085-3 | Sequence 3, Appli |
| 10 | 44 | 43.6 | 49377 | 1 US-08-764-233A-1 | Sequence 1, Appli |
| 11 | 41.2 | 40.8 | 43280 | 2 US-08-804-227C-1 | Sequence 1, Appli |
| 12 | 38.8 | 38.4 | 13987 | 2 US-08-804-227C-13 | Sequence 13, Appli |
| 13 | 38.8 | 38.4 | 44377 | 2 US-08-804-227C-7 | Sequence 7, Appli |
| 14 | 38.8 | 38.4 | 44377 | 2 US-08-804-198-1 | Sequence 1, Appli |
| 15 | 38.8 | 38.4 | 80161 | 3 US-09-036-987A-1 | Sequence 1, Appli |
| 16 | 38.8 | 38.4 | 80161 | 4 US-09-370-700-1 | Sequence 1, Appli |
| 17 | 37 | 36.6 | 13842 | 4 US-09-105-537-30 | Sequence 30, Appli |
| 18 | 37 | 36.6 | 36778 | 4 US-09-105-537-5 | Sequence 5, Appli |
| 19 | 37 | 36.6 | 38506 | 3 US-09-320-878-19 | Sequence 19, Appli |
| 20 | 35 | 34.7 | 28958 | 1 US-08-258-261B-6 | Sequence 6, Appli |
| 21 | 35 | 34.7 | 28958 | 1 US-08-456-837-6 | Sequence 6, Appli |
| 22 | 35 | 34.7 | 28958 | 1 US-08-457-342-6 | Sequence 6, Appli |
| 23 | 35 | 34.7 | 28958 | 1 US-08-457-646A-6 | Sequence 6, Appli |
| 24 | 35 | 34.7 | 28958 | 1 US-08-458-076A-6 | Sequence 6, Appli |
| 25 | 35 | 34.7 | 28958 | 1 US-08-764-233A-4 | Sequence 4, Appli |
| 26 | 35 | 34.7 | 28958 | 1 US-08-457-335A-6 | Sequence 6, Appli |
| 27 | 35 | 34.7 | 28958 | 1 US-08-729-214-6 | Sequence 6, Appli |

| | | | | | | | |
|----|------|------|-------|---------|------------------|--------------------|--------------------|
| 28 | 35 | 34.7 | 28958 | 3 | US-09-028-934-6 | Sequence 6, Appli | |
| 29 | 32.4 | 32.1 | 50937 | 4 | US-09-428-517-1 | Sequence 1, Appli | |
| c | 30 | 32.4 | 32.1 | 4403765 | 4 | US-09-103-840A-2 | Sequence 2, Appli |
| 31 | 30.8 | 30.5 | 11219 | 1 | US-07-642-734C-1 | Sequence 1, Appli | |
| 32 | 30.8 | 30.5 | 11219 | 3 | US-08-439-009A-1 | Sequence 1, Appli | |
| 33 | 28.4 | 28.1 | 2493 | 1 | US-07-977-434-5 | Sequence 5, Appli | |
| 34 | 28.4 | 28.1 | 2493 | 1 | US-08-458-819-5 | Sequence 5, Appli | |
| 35 | 28.4 | 28.1 | 2493 | 5 | PCT-US91-07035-5 | Sequence 5, Appli | |
| 36 | 28.2 | 27.9 | 70000 | 4 | US-09-851-896-3 | Sequence 3, Appli | |
| 37 | 27.4 | 27.1 | 2634 | 3 | US-08-911-853-30 | Sequence 30, Appli | |
| 38 | 27.4 | 27.1 | 2634 | 4 | US-09-479-409-30 | Sequence 30, Appli | |
| 39 | 27.4 | 27.1 | 2634 | 4 | US-09-479-453-30 | Sequence 29, Appli | |
| 40 | 27.4 | 27.1 | 17612 | 3 | US-08-911-853-29 | Sequence 29, Appli | |
| 41 | 27.4 | 27.1 | 17612 | 4 | US-09-479-409-29 | Sequence 29, Appli | |
| 42 | 27.4 | 27.1 | 17612 | 4 | US-09-479-453-29 | Sequence 29, Appli | |
| c | 43 | 26.8 | 26.5 | 1811 | 4 | US-09-280-116-94 | Sequence 94, Appli |
| c | 44 | 26.8 | 26.5 | 2085 | 3 | US-09-299-843A-65 | Sequence 65, Appli |
| c | 45 | 26.8 | 26.5 | 2085 | 4 | US-09-088-337B-65 | Sequence 65, Appli |

ALIGNMENTS

```
RESULT 1
US-09-335-409-1
Sequence 1, Application US/09335409
Patent No. 6121029
GENERAL INFORMATION:
APPLICANT: Schupp, Thomas
APPLICANT: Ligon, James
APPLICANT: Molnar, Istvan
APPLICANT: Zirkle, Ross
APPLICANT: Cyr, Devon
TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
FILE REFERENCE: 4-30582A
CURRENT APPLICATION NUMBER: US/09/335.409
CURRENT FILING DATE: 1999-06-17
NUMBER OF SEQ ID NOS: 30
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 68750
TYPE: DNA
ORGANISM: Sorangium cellulosum
US-09-335-409-1

Query Match      100.0%; Score 101; DB 3; Length 68750;
Best Local Similarity 100.0%; Pred. No. 2.7e-23;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      1 CGCGGAGTGGCGCTTACAGGCGGTGATCATGACAGCCGAGCGCTCATGATGATGTGTC 60
Db      36162 CGGCGAGATGGCGCTTACAGGCGGTGATCATGACAGCCGAGCGCTCATGATGATGTGTC 36221
Oy      61 TTGATGAGCAGACGACGACGCGCTTCTCGGCGGTCTGTGGCA 101
Db      36222 TTGATGAGCAGACGACGACGCGCTTCTCGGCGGTCTGTGGCA 36262

RESULT 2
US-09-568-102-1
Sequence 1, Application US/09568102
Patent No. 6346404
GENERAL INFORMATION:
APPLICANT: Schupp, Thomas
APPLICANT: Ligon, James
APPLICANT: Molnar, Istvan
APPLICANT: Zirkle, Ross
APPLICANT: Cyr, Devon
TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
FILE REFERENCE: 4-30582A
CURRENT APPLICATION NUMBER: US/09/568.102
```

```
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-102-1
```

```
Query Match          100.0%; Score 101; DB 4; Length 68750;
Best Local Similarity 100.0%; Pred. No. 2.7e-23;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      1 CGGCGGAGTGCCGCTTACAGAGGCGGTGATCGACGCCGAGCGCTCGATGATGTTGTC 60
          |||||||
Db 36162 CGGCGGAGTGCCGCTTACAGAGGCGGTGATCCATCGACGCCGAGCGCTCGATGATGTTGTC 36221
```

```
OY      61 TTGATGAGCAGACCAACCGACCGCTTCTCGCGGGTGTGGCA 101
          |||||||
Db 36222 TTGATGAGCAGACCAACCGACCGCTTCTCGCGGGTGTGGCA 36262
```

```
RESULT 3
US-09-567-969-1
; Sequence 1, Application US/09567969
; Patent No. 6355457
; GENERAL INFORMATION:
```

```
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/567,969
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-567-969-1
```

```
Query Match          100.0%; Score 101; DB 4; Length 68750;
Best Local Similarity 100.0%; Pred. No. 2.7e-23;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      1 CGGCGGAGTGCCGCTTACAGAGGCGGTGATCGACGCCGAGCGCTCGATGATGTTGTC 60
          |||||||
Db 36162 CGGCGGAGTGCCGCTTACAGAGGCGGTGATCCATCGACGCCGAGCGCTCGATGATGTTGTC 36221
```

```
OY      61 TTGATGAGCAGACCAACCGACCGCTTCTCGCGGGTGTGGCA 101
          |||||||
Db 36222 TTGATGAGCAGACCAACCGACCGCTTCTCGCGGGTGTGGCA 36262
```

```
RESULT 4
US-09-568-480-1
; Sequence 1, Application US/09568480
; Patent No. 6355458
; GENERAL INFORMATION:
```

```
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
```

```
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,480
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-480-1
```

```
Query Match          100.0%; Score 101; DB 4; Length 68750;
Best Local Similarity 100.0%; Pred. No. 2.7e-23;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      1 CGGCGGAGTGCCGCTTACAGAGGCGGTGATCCATCGACGCCGAGCGCTCGATGATGTTGTC 60
          |||||||
Db 36162 CGGCGGAGTGCCGCTTACAGAGGCGGTGATCCATCGACGCCGAGCGCTCGATGATGTTGTC 36221
```

```
OY      61 TTGATGAGCAGACCAACCGACCGCTTCTCGCGGGTGTGGCA 101
          |||||||
Db 36222 TTGATGAGCAGACCAACCGACCGCTTCTCGCGGGTGTGGCA 36262
```

```
RESULT 5
US-09-568-486-1
; Sequence 1, Application US/09568486
; Patent No. 6355459
; GENERAL INFORMATION:
```

```
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,486
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-486-1
```

```
Query Match          100.0%; Score 101; DB 4; Length 68750;
Best Local Similarity 100.0%; Pred. No. 2.7e-23;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      1 CGGCGGAGTGCCGCTTACAGAGGCGGTGATCCATCGACGCCGAGCGCTCGATGATGTTGTC 60
          |||||||
Db 36162 CGGCGGAGTGCCGCTTACAGAGGCGGTGATCCATCGACGCCGAGCGCTCGATGATGTTGTC 36221
```

```
OY      61 TTGATGAGCAGACCAACCGACCGCTTCTCGCGGGTGTGGCA 101
          |||||||
Db 36222 TTGATGAGCAGACCAACCGACCGCTTCTCGCGGGTGTGGCA 36262
```

```
RESULT 6
US-09-568-472-1
; Sequence 1, Application US/09568472
; Patent No. 6358719
; GENERAL INFORMATION:
```

```
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
```

```

: APPLICANT: Zirkle, Ross
: APPLICANT: Cyr, Devon
: APPLICANT: Goerlach, Joern
: TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
: FILE REFERENCE: 4-30582A
: CURRENT APPLICATION NUMBER: 09/335,409
: CURRENT FILING DATE: 2000-05-10
: PRIOR APPLICATION NUMBER: 09/335,409
: PRIOR FILING DATE: 1999-06-17
: NUMBER OF SEQ ID NOS: 30
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 1
: LENGTH: 68750
: TYPE: DNA
: ORGANISM: Sorangium cellulosum
US-09-568-472-1

```

| | | | | |
|---------------------------|--------|--------------------|-------|---------------|
| Query Match | 100.0% | Score 101; | DB 4; | Length 66750; |
| Best Local Similarity | 100.0% | Pred. No. 2.7e-23; | | |
| Matches 101; Conservative | 0; | Mismatches | 0; | Gaps 0; |

Qy 1 CGCGGAGTGGCCGTACAGGGCGTGATCCATGCAGCCGAGCGCTCATGATGGTGTGC 60
|||||
|||||
|||||
Db 36162 CGCGGAGTGGCCGTACAGGGCGTGATCCATGCAGCCGAGCGCTCATGATGGTGTGC 36221

61 TTGATGAGCAGACCACCGCTTCTCGCGGTGCTGCA 101

Db 36222 TTGATGAGCAGACCACCGCGTCTCGGGTGCTGGCA 36262

RESULT 7
US-09-567-899-1

GENERAL INFORMATION:
 PATENT NO: 30582A
 APPLICANT: Schnupp, Thomas
 APPLICANT: Ligon, James
 APPLICANT: Molnar, Istvan
 APPLICANT: Zirkle, Ross
 APPLICANT: Cyr, Devon
 APPLICANT: Goerlach, Joern
 TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
 FILE REFERENCE: 4-30582A

```

: CURRENT FILING DATE: 2000-05-10
: PRIOR APPLICATION NUMBER: 09/335,409
: PRIOR FILING DATE: 1999-06-17
: NUMBER OF SEQ ID NOS: 30
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 1

```

```
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-567-899-1
```

| | | | | |
|---------------------------|--------|--------------------|-------|-------------------|
| Query Match | 100.0% | Score 101; | DB 4; | Length 68750; |
| Best Local Similarity | 100.0% | Pred. No. 2.7e-23; | | |
| Matches 101; Conservative | 0; | Mismatches | 0; | Indels 0; Gaps 0; |

Qy 1 CGGGCGAGTGGCCGTTACAGGCGGTGATCCATGCAGCCGAGCGCTCGATGATGGTGTGC 60
|||||
Db 36162 CGGCGGAGTGGCGTTACAGGCGGTGATCCATGCAGCCGAGCGCTCGATGATGGTGTGC 36227

61 TTGATGAGCAGACCACCGACTTCTCGCGGTGCTGCA 101

Db 36222 TTGATGAGCAGACCACCGCTTCTCGCGGGTGCTGGCA 36262

```

RESULT 8
US-09-443-501A-2
; Sequence 2, Application US/09443501A
; Patent No. 6303342
; GENERAL INFORMATION:

```

APPLICANT: Kosan Biosciences, Inc.

APPLICANT: Katz, Leonard

APPLICANT: Tang, Li

TITLE OF INVENTION:

FILE REFERENCE: 30062-20031.00
CURRENT ADDITION NUMBER: ITS 600 443 501A

CURRENT FILING DATE: 1999-11-19
PRIOR APPLICATION NUMBER: PG 60/130 660

PRIOR FILING DATE: 1999-04-22

PRIOR FILING DATE: 1999-03-03
PRIOR APPLICATION NUMBER: US 60/110 386

PRIOR FILING DATE: 1999-02-10
PRIOR APPLICATION NUMBER: US 60/109 401

PRIOR FILING DATE: 1998-11-20
NUMBER OF SEQ ID NOS: 33

```

; SOFTWARE: Fastseq for windows version 4.0
; SEQ ID NO 2

```

LENGTH: 71985
TYPE: DNA

ORGANISM: Artificial Sequence

OTHER INFORMATION: Synthetic construct
HS-08-443-501A-3

US-09-443-501A-2

| | | | | |
|---------------------------|---------|--------------------|-------|---------------|
| Query Match | 100.0%; | Score 101; | DB 4; | Length 71989; |
| Best Local Similarity | 100.0%; | Pred. No. 2.7e-23; | | |
| Matches 101; Conservative | 0; | Mismatches | 0; | Gaps 0; |

QY 1 CGGCGGAGTGGCCGTTACAGGCGGTGATCCATGACCGGAGCCCTCGATGATGGTGTGC 60
|||||
Db 30550 CGGCGGAGTGGCCGTTACAGGCGGTGATCCATGACCGGAGCCCTCGATGATGGTGTGC 30609

61 TTGATGAGCAGACCACCGCTTCTCGGGGTCTGGCA 101
QY

Db 30610 TTGATGAGCAGACCACCGCTTCTCGCGGGTGCTGCA 30650

RESULT 9
US-09-144-085-3

; Patent No. 6280999

APPLICANT: Gustafsson, Claes

APPLICANT: Ashley, Gary

APPLICANT: Ziermann, Rainer
TITLE OF INVENTION: CODING AND ENCODING DNA

1. TITLE OF INVENTION: THEREFOR
2. REFERENCE: 20063-20030 20

CURRENT APPLICATION NUMBER: US/09/144,085
CURRENT FILING DATE: 10-08-08-21

EARLIER APPLICATION NUMBER: 09/010,809

NUMBER OF SEQ ID NOS: 8

```

; SEQ ID NO 3
          LENGTH: 33530

```

```

; TYPE: DNA
; OPG&NTSM, GOR

```

US-09-144-085-3

| | Query Match | 51.1% | Score 51.6 | DB 4 | Length 33529 |
|----|--|----------------|-------------------|----------|--------------|
| | Best Local Similarity | 70.4% | Pred. No. 1.5e-07 | | |
| | Matches 69 | Conservative 0 | Mismatches 29 | Indels 0 | Gaps 0 |
| QY | 3 GCGAGTGGCCGTTACAGGCGCTATTCATGAGCGGAGCCCTCGAGTATGCGTGT 62 | | | | |
| | | | | | |

Db 14155 GCGGCGCCGCTGACGCGGCTCTGACCTGCGCGCGCTGACGAGCGGCTGCTC 14214
Qy 63 GATGACGACGACGACGACGCTTCTCGCGGCTGCTGCG 100
Db 14215 GCGGCGCAGACGCGGCGGCTCTCGCGGCTGCTGCG 14252

RESULT 10
US-08-764-233A-1
; Sequence 1, Application US/08764233A
; Patent No. 5716849

GENERAL INFORMATION:
APPLICANT: Ligon, James M.
APPLICANT: Schupp, Thomas
APPLICANT: Beck, James J.
APPLICANT: Hill, Dwight S.
APPLICANT: Neff, Snezana
APPLICANT: Ryals, John A.
TITLE OF INVENTION: Genes for The Biosynthesis Of Soraphen
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ciba-Geigy Corporation
STREET: 520 White Plains Road, P.O. Box 2005
CITY: Tarrytown
STATE: NY
COUNTRY: USA
ZIP: 10591

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/764,233A
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/729,214
FILING DATE: 09-OCT-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/258,261
FILING DATE: 08-JUN-1994
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: 1506/C1P6
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919) 541-8587
TELEFAX: (919) 541-8689
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 49377 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Sorangium cellulosum
IMMEDIATE SOURCE:
CLONE: p98/L, pL3, and pVKM15

FEATURE:
NAME/KEY: misc_feature
LOCATION: 383..760
OTHER INFORMATION: /product= "Sora"
OTHER INFORMATION: /note= "this gene encodes a protein that is highly homologous
OTHER INFORMATION: the reductase domains of type I PKs such as eryA from
OTHER INFORMATION: Saccharopolyspora erythraea."
FEATURE:
NAME/KEY: misc_feature
LOCATION: 927..19874
OTHER INFORMATION: /product= "Sora"
OTHER INFORMATION: /note= "Gene product is highly homologous to type I PKs that
OTHER INFORMATION: are known to be involved in the synthesis of polyketide

OTHER INFORMATION: compounds."

FEATURE:
NAME/KEY: misc_feature
LOCATION: 942..7115
OTHER INFORMATION: /product= "Module 1 of Sora"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 7203..12884
OTHER INFORMATION: /product= "Module 2 of Sora"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 13455..19616
OTHER INFORMATION: /product= "Module 3 of Sora"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 19871..46318
OTHER INFORMATION: /product= "Sora"
OTHER INFORMATION: /note= "Gene product is highly homologous to type I PKs ge
FEATURE:
NAME/KEY: misc_feature
LOCATION: 19870..24556
OTHER INFORMATION: /product= "Module 1 of Sora"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 24638..30820
OTHER INFORMATION: /product= "Module 2 of Sora"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 30881..35446
OTHER INFORMATION: /product= "Module 3 of Sora"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 35528..40114
OTHER INFORMATION: /product= "Module 4 of Sora"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 40190..46318
OTHER INFORMATION: /product= "Module 5 of Sora"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 46851..47891
OTHER INFORMATION: /product= "Sora"
OTHER INFORMATION: /note= "The protein encoded by the sora gene is highly
OTHER INFORMATION: homologous to the methyltransferase from Streptomyces
OTHER INFORMATION: hygroscopicus that is involved in the synthesis of the
OTHER INFORMATION: polyketide rapamycin."

US-08-764-233A-1

Query Match 43.6%; Score 44; DB 1; Length 49377;
Best Local Similarity 65.0%; Pred. No. 4.5e-05;
Matches 65; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

Qy 1 GCGGCGAGTGGCCGTTACAGAGGGGTGATCATCACCACCGAGGCTGATGATGCTGCG 60
Db 6277 CGTCGACGCTGCCCGCTACGACGCGGTCTTCACAGCGAGAGCTCGCAGATGCGCTGA 6336

Qy 61 TTGATGACGACGACGACGACGCTTCTCGCGGCTGCTGCG 100
Db 6337 TCGGCGACATGACGCGCGGAGCGCATGAGCGGGTCTTGC 6376

RESULT 11
US-08-804-227C-1
; Sequence 1, Application US/08804227C
; Patent No. 5876991

GENERAL INFORMATION:
APPLICANT: Dehoff, Bradley S.
APPLICANT: Kuhnstoss, Stuart A.
APPLICANT: Rostock, Paul R., Jr.
APPLICANT: Sutton, Kimberly L.
TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:

```

: ADDRESSEE: THOMAS G. PLANT 1501
: STREET: LILLY CORPORATE CENTER
: CITY: INDIANAPOLIS
: STATE: IN
: COUNTRY: USA
: ZIP: 46285
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: MS-DOS
: SOFTWARE: ASCII(DOS) Text only
:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/804,227C
: FILING DATE: February 21, 1997
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Plant, Thomas, G.
: REGISTRATION NUMBER: 35,784
: REFERENCE/DOCKET NUMBER: X-8231
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 317-276-2459
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 43280 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 816..14234
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 14351..19945
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 20010..31199
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 31232..36067
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 36249..41774
:
: US-08-804-227C-1
:
: Query Match 40.8%; Score 41.2; DB 2; Length 43280;
: Best Local Similarity 67.4%; Pred. No. 0.00035;
: Matches 58; Conservative 0; Mismatches 28; Indels 0; Gaps 0;
:
: QY 1 CGGGGAGTGGCGCTTACAGGCGGTGATCGACCGGAGCGCTCGATGATGCTGTC 60
: DB 13093 CGGGCGACGGCGCGCTGACCGGAGTGATCGACGCGCGGCGGTCTCGACGACGACAC 13152
:
: QY 61 TTGATGACAGACACCGACCGCTTC 86
: DB 13153 TCGACGGCGCTGACCCCGGAGCGATC 13178
:
: RESULT 12
: US-08-804-227C-13
: Sequence 13, Application US/08804227C
: Patent No. 5876991
:
: GENERAL INFORMATION:
: APPLICANT: Dehoff, Bradley S.
: APPLICANT: Kuhstoss, Stuart A.
: APPLICANT: Rostock, Paul R., Jr.
: APPLICANT: Sutton, Kimberly L.
: TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
: NUMBER OF SEQUENCES: 15
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: THOMAS G. PLANT 1501
: STREET: LILLY CORPORATE CENTER
: CITY: INDIANAPOLIS

```

```

: STATE: IN
: COUNTRY: USA
: ZIP: 46285
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: MS-DOS
: SOFTWARE: ASCII(DOS) Text only
:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/804,227C
: FILING DATE: February 21, 1997
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Plant, Thomas, G.
: REGISTRATION NUMBER: 35,784
: REFERENCE/DOCKET NUMBER: X-8231
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 317-276-2459
: INFORMATION FOR SEQ ID NO: 13:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 13987 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 350..13987
:
: US-08-804-227C-13
:
: Query Match 38.4%; Score 38.8; DB 2; Length 13987;
: Best Local Similarity 62.2%; Pred. No. 0.0016;
: Matches 61; Conservative 0; Mismatches 37; Indels 0; Gaps 0;
:
: QY 1 CGGGGAGTGGCGCTTACAGGCGGTGATCGACCGGAGCGCTCGATGATGCTGTC 60
: DB 7260 CCGAGAGCGGCGCGCTGCTCCCGTACTCGACGCGGCGGTGCTGTCGACGACGCTGTC 7319
:
: QY 61 TTGATGACAGACACCGACCGCTTCCTGCGGCTGCTG 98
: DB 7320 TCGACTCGCTGACCTCGACCGGCTGAGCGCGCTGCTG 7357
:
: RESULT 13
: US-08-804-227C-7
: Sequence 7, Application US/08804227C
: Patent No. 5876991
:
: GENERAL INFORMATION:
: APPLICANT: Dehoff, Bradley S.
: APPLICANT: Kuhstoss, Stuart A.
: APPLICANT: Rostock, Paul R., Jr.
: APPLICANT: Sutton, Kimberly L.
: TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
: NUMBER OF SEQUENCES: 15
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: THOMAS G. PLANT 1501
: STREET: LILLY CORPORATE CENTER
: CITY: INDIANAPOLIS
: STATE: IN
: COUNTRY: USA
: ZIP: 46285
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: MS-DOS
: SOFTWARE: ASCII(DOS) Text only
:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/804,227C
: FILING DATE: February 21, 1997
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Plant, Thomas, G.
: REGISTRATION NUMBER: 35,784

```

REFERENCE/DOCKET NUMBER: X-8231
TELECOMMUNICATION INFORMATION:
TELEPHONE: 317-276-2459
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 4437 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 350..14002
FEATURE:
NAME/KEY: CDS
LOCATION: 14046..20036
FEATURE:
NAME/KEY: CDS
LOCATION: 20110..31284
FEATURE:
NAME/KEY: CDS
LOCATION: 31329..36071
FEATURE:
NAME/KEY: CDS
LOCATION: 36155..41830
US-08-804-227C-7

Query Match Best Local Similarity 62.2%; Score 38.8; DB 2; Length 44377;
Matches 61; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 1 CGCGGAGTGGCCGTTACAGAGCGGTATCCATCCAGCCGAGCGCTCGATGATGTGTGC 60
DB 7275 CCAGAGAGGGGCCCTCGTGGCGTACTGCACGCGGAGGTGTGTCGAGAGGGTGTGC 7334
QY 61 TTGATGACGACGACCGACCGACCGCTTCTCGCGGTGCTG 98
DB 7335 TCGACTGCTCACCTCCGACCGGGTGGACCGCCGCTACTG 7372

RESULT 14

US-08-804-198-1
Sequence 1, Application US/08804198
Patent No. 5945320
GENERAL INFORMATION:
APPLICANT: Burgett, Stanley G.
APPLICANT: Kuhstoss, Stuart A.
APPLICANT: Rao, Nagaraja R.
APPLICANT: Richardson, Mark A.
APPLICANT: Rosteck, Paul R., Jr.
TITLE OF INVENTION: PLATENOLIDE SYNTHASE GENE
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: PAUL R. CANTRELL 1138
STREET: LILLY CORPORATE CENTER
CITY: INDIANAPOLIS
STATE: IN
COUNTRY: USA
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Macintosh
OPERATING SYSTEM: Macintosh 7.0
SOFTWARE: Microsoft Word 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/804,198
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: CANTRELL, PAUL R.
REGISTRATION NUMBER: 36,470
REFERENCE/DOCKET NUMBER: P9113
TELECOMMUNICATION INFORMATION:

TELEPHONE: 317-276-3885
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4437 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 350..14002
FEATURE:
NAME/KEY: CDS
LOCATION: 14046..20036
FEATURE:
NAME/KEY: CDS
LOCATION: 20110..31284
FEATURE:
NAME/KEY: CDS
LOCATION: 31329..36071
FEATURE:
NAME/KEY: CDS
LOCATION: 36155..41830
US-08-804-198-1

Query Match Best Local Similarity 62.2%; Score 38.8; DB 2; Length 44377;
Matches 61; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 1 CGCGGAGTGGCCGTTACAGAGCGGTATCCATCCAGCCGAGCGCTCGATGATGTGTGC 60
DB 7275 CCAGAGAGGGGCCCTCGTGGCGTACTGCACGCGGAGGTGTGTCGAGAGGGTGTGC 7334
QY 61 TTGATGACGACGACCGACCGACCGCTTCTCGCGGTGCTG 98
DB 7335 TCGACTGCTCACCTCCGACCGGGTGGACCGCCGCTACTG 7372

RESULT 15

US-09-036-987A-1
Sequence 1, Application US/09036987A
Patent No. 6143526
GENERAL INFORMATION:
APPLICANT: Baltz, Richard H.
APPLICANT: Crawford, Mary C.
APPLICANT: Crowford, Kathryn P.
APPLICANT: Madduri, Krishnamurthy
APPLICANT: Merlo, Donald J.
APPLICANT: Treadway, Patli J.
APPLICANT: Turner, Jan R.
TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
NUMBER OF SEQUENCES: 39
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dow Agrosciences LLC Patent Department
STREET: 9330 Zionsville Road
CITY: Indianapolis
STATE: Indiana
COUNTRY: USA
ZIP: 46268
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/036,987A
FILING DATE: 09-MAR-1998
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Stuart, Donald R
REGISTRATION NUMBER: 28,479

```

; REFERENCE/DOCKET NUMBER: 50,608
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (317)337-4816
; TELEFAX: (317)337-4847
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 80161 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-09-036-987A-1

```

```

Query Match 38.4%; Score 38.8; DB 3; Length 80161;
Best Local Similarity 62.2%; Pred. No. 0.0023;
Matches 61; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

```

```

QY 1 CGGCGAGTGGCGTTTACAGGGCGTGATCATGACAGCCGAGCGCTCGATGATGATGTC 60
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 74828 CTGGGGAACCCGTTGGGGTGGTGTGCACGCCCGCGTGTGCTGATGACGGTGTGC 74887
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 61 TTGATGAGCAGACACGACCGCTTCTCGCGGGTCTG 98
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 74888 TGATGTCATGTGCGCGGAGCGCTTGACGCGGTGTG 74925
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

Search completed: November 6, 2002, 15:59:48
 Job time : 201.625 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 6, 2002, 11:51:49 ; Search time 157 Seconds
(without alignments)
1448.738 Million cell updates/sec

Title: US-09-724-876-2_COPY_30550_30650

Perfect score: 101

Sequence: 1 cggcgagagcgccgttacg.....gtcttcgagcggtcgtagc.101

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

N_Geneseq_101002.*
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1 | 101 | 100.0 | 68750 | 21 | AAZ55887 |
| 2 | 101 | 100.0 | 71989 | 21 | AAZ29349 |
| 3 | 51.6 | 51.1 | 33529 | 23 | AAZ17367 |
| 4 | 46.8 | 46.3 | 31422 | 21 | AAZ92302 |
| 5 | 46.8 | 46.3 | 31422 | 22 | AAH79278 |
| 6 | 44.4 | 43.6 | 49377 | 19 | AAV05287 |
| 7 | 41.2 | 40.8 | 43280 | 18 | AAH80413 |
| 8 | 41.2 | 40.8 | 125401 | 22 | AAH17186 |
| 9 | 40.4 | 40.0 | 5088 | 22 | AAZ90038 |

| | | | | | | |
|----|------|------|---------|----|-----------|-----------------------|
| 10 | 40.4 | 40.0 | 20394 | 22 | AAZ24892 | Pimaricin biosynth |
| 11 | 40.4 | 40.0 | 34071 | 22 | AAZ90033 | Nucleotide sequenc |
| 12 | 40.4 | 40.0 | 42717 | 22 | AAZ90032 | Nucleotide sequenc |
| 13 | 39.4 | 39.0 | 4851 | 22 | AAZ52052 | Mycobacterium tube |
| 14 | 39.2 | 38.8 | 552 | 24 | AAI68063 | Epochillone PKS KR |
| 15 | 39.2 | 38.8 | 552 | 24 | AAI68064 | 9-keto-epochillone |
| 16 | 38.8 | 38.4 | 13987 | 18 | AAH80415 | Hydrl srmf/tylg O |
| 17 | 38.8 | 38.4 | 16767 | 22 | AAZ83339 | S. spinoosa DNA fra |
| 18 | 38.8 | 38.4 | 29736 | 22 | AAZ83317 | S. spinoosa DNA fra |
| 19 | 38.8 | 38.4 | 44377 | 18 | AAZ78508 | Platenolide synth |
| 20 | 38.8 | 38.4 | 44377 | 18 | AAZ80414 | Platenolide synth |
| 21 | 38.8 | 38.4 | 50000 | 22 | AAZ88313 | S. spinoosa DNA fra |
| 22 | 38.8 | 38.4 | 53789 | 19 | AAZ21187 | Amycolatopsis medl |
| 23 | 38.8 | 38.4 | 80161 | 20 | AAZ21501 | DNA fragment of Sa |
| 24 | 38.2 | 37.8 | 50000 | 22 | AAZ88316 | S. spinoosa DNA fra |
| 25 | 37.2 | 36.8 | 2700 | 22 | AAZ17193 | ERD48 insert DNA i |
| 26 | 37.2 | 36.8 | 65140 | 22 | AAZ17184 | Streptomyces nous |
| 27 | 37 | 36.6 | 13842 | 21 | AAZ87297 | S. venezuelae macr |
| 28 | 37 | 36.6 | 36778 | 21 | AAZ87318 | S. venezuelae pik |
| 29 | 37 | 36.6 | 37948 | 21 | AAZ87285 | S. venezuelae pik |
| 30 | 37 | 36.6 | 38505 | 21 | AAZ75633 | Nucleotide sequenc |
| 31 | 37 | 36.6 | 38506 | 21 | AAZ56001 | Recombinant cosmid |
| 32 | 36 | 35.6 | 5292 | 22 | AAZ90036 | Nucleotide sequenc |
| 33 | 35 | 34.7 | 27541 | 22 | AAZ17185 | Streptomyces nous |
| 34 | 35 | 34.7 | 28598 | 17 | AAZ06769 | Soraniangium cellulos |
| 35 | 35 | 34.7 | 28958 | 18 | AAZ89596 | Soraniangium cellulos |
| 36 | 35 | 34.7 | 28958 | 21 | AAZ75299 | DNA sequence of 50 |
| 37 | 34.6 | 34.3 | 77536 | 21 | AAZ14651 | Nucleotide sequenc |
| 38 | 34.2 | 32.1 | 77536 | 21 | AAZ14651 | Nucleotide sequenc |
| 39 | 32.4 | 32.1 | 6459 | 22 | AAZ88336 | S. spinoosa DNA fra |
| 40 | 32.4 | 32.1 | 50937 | 21 | AAZ09469 | Streptococcus olea |
| 41 | 32.4 | 32.1 | 4403765 | 22 | AAZ19683 | Mycobacterium tube |
| 42 | 32.2 | 31.9 | 8301 | 22 | AAZ90035 | Nucleotide sequenc |
| 43 | 30.8 | 30.5 | 29879 | 22 | AAZ046806 | erya region of S. |
| 44 | 30.6 | 30.3 | 1371 | 14 | AAZ52066 | Mycobacterium tube |
| 45 | 30.6 | 30.3 | 1371 | 22 | AAZ97009 | Mycobacterium tube |

ALIGNMENTS

| | | |
|----------|--|--------------------------|
| RESULT 1 | AAZ55887 | standard; DNA; 68750 BP. |
| ID | AAZ55887 | |
| AC | AAZ55887 | |
| XX | 10-APR-2000 | (first entry) |
| DT | | |
| XX | | |
| DE | Soraniangium cellulosum 68.75 kb contig. | |
| XX | | |
| KW | Epochillone biosynthesis; type I polyketide synthase; taxol substitue; | |
| KX | anticancer; ds. | |
| XX | | |
| OS | Soraniangium cellulosum. | |
| XX | | |
| EH | Location/Qualifiers | |
| FT | 1..1826 | |
| FT | /*tag= a | |
| FT | /partial | |
| FT | /product= "Partial Orf 1 protein (AAZ55880)" | |
| FT | /note= "No initiation codon given in the specification" | |
| FT | complement (1900..3171) | |
| FT | /*tag= b | |
| FT | /product= "Orf 2 protein (AAZ55881)" | |
| FT | 3415..5556 | |
| FT | /*tag= c | |
| FT | /product= "Orf 3 protein (AAZ55882)" | |
| FT | complement (5612..55992) | |
| FT | /*tag= d | |
| FT | /product= "Orf 4 protein (AAZ55883)" | |
| FT | 6226..6675 | |
| FT | /*tag= e | |
| FT | CDS | |

```
FT      /product= "Orf 5 protein (AAV58584)"
FT      /tag= f
FT      /product= "Type I polyketide synthase, EPOS A
FT      11872..116104
FT      /tag= g
FT      /product= "Non-ribosomal peptide synthetase, EPOS P
FT      (AAV58574)"
FT      16251..21749
FT      /tag= h
FT      /product= "Type I polyketide synthase, EPOS B
FT      (AAV58575)"
FT      21746..43519
FT      /tag= i
FT      /product= "Type I polyketide synthase, EPOS C
FT      (AAV58576)"
FT      43524..54920
FT      /tag= j
FT      /product= "Type I polyketide synthase, EPOS D
FT      (AAV58577)"
FT      54935..62254
FT      /tag= k
FT      /product= "Type I polyketide synthase, EPOS E
FT      (AAV58578)"
FT      62369..63628
FT      /tag= l
FT      /product= "Cytochrome P450 oxygenase homologue, EPOS F
FT      (AAV58579)"
FT      63779..64333
FT      /tag= m
FT      /product= "Orf 6 protein (AAV58585)"
FT      complement (63853..64290)
FT      /tag= n
FT      /product= "Orf 7 protein (AAV58586)"
FT      64363..64920
FT      /tag= o
FT      /product= "Orf 8 protein (AAV58587)"
FT      complement (64287..64727)
FT      /tag= p
FT      /product= "Orf 9 protein (AAV58588)"
FT      65063..65767
FT      /tag= q
FT      /product= "Orf 10 protein"
FT      complement (65008..65874)
FT      /tag= r
FT      /product= "Orf 11 protein (AAV58590)"
FT      complement (65871..66338)
FT      /tag= s
FT      /product= "Orf 12 protein (AAV58591)"
FT      66667..67137
FT      /tag= t
FT      /product= "Orf 13 protein (AAV58592)"
FT      67334..68251
FT      /tag= u
FT      /product= "Orf 14 protein (AAV58593)"
FT      68346..68750
FT      /tag= v
FT      /product= "Partial Orf 15 protein (AAV58594)"
FT      /note= "No termination codon given in the specification"
XX      W09966028-A2.
XX      PD
XX      23-DEC-1999.
XX      PF
XX      16-JUN-1999; 99WO-EP04171.
XX      PR
XX      18-JUN-1998; 98US-0099504.
XX      24-SEP-1998; 98US-0101631.
XX      PR
XX      05-FEB-1999; 99US-0118906.
XX      PA
XX      (NOVS ) NOVARTIS AG.
```

```
PA      (NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.
XX
XX      Schupp T, Ligon JM, Molnar I, Zirkle R, Goerlach J, Cyr D:
XX      WPI, 2000-097741/08.
DR      P-PSDB: AAV58573, AAV58574, AAV58575, AAV58576, AAV58577, AAV58578,
DR      AAV58579, AAV58580, AAV58581, AAV58582, AAV58583, AAV58584,
DR      AAV58585, AAV58586, AAV58587, AAV58588, AAV58590, AAV58591,
DR      AAV58592, AAV58593, AAV58594.
XX
PT      New isolated epoethione synthase genes, used for the recombinant
PT      production of epoethione for use in cancer therapy .
XX
XX      Claim 14; Page 87-104; 174pp; English.
XX
CC      This sequence represents a 68.75 kb contig from Sorangium cellulosum
CC      comprising 22 open reading frames (ORFs) and includes genes encoding
CC      proteins involved in the biosynthesis of epoethiones. Epoethiones A and
CC      B are 16-membered macrocyclic polyketides with an acylcysteine-derived
CC      starter unit; polyketides being synthesised from two-carbon building
CC      blocks, the beta-carbon of which always carries a keto group. Each round
CC      of two-carbon addition is carried out by a complex of enzymes known as
CC      the polyketide synthase in a manner similar to fatty acid biosynthesis.
CC      EPOS A (AAV58573) and EPOS P (AAV58574) are involved in formation of
CC      the thiazole ring formation of epoethiones, and EPOS B, EPOS C, EPOS D
CC      and EPOS E (AAV58575-Y58578) are involved in polyketide backbone
CC      formation. EPO F (AAV58579) is an epoethione macrolactone oxidase, and
CC      the proteins Orf 3 (AAV58582) and Orf14 (AAV58593) are thought to be
CC      involved in transport. Epoethiones mimic the biological activity of
CC      taxol, and may be substituted for taxol in cancer chemotherapeutic
CC      compositions. Epoethiones exhibit a much lower drop in potency against a
CC      multiply drug-resistant cell line compared with taxol, and are
CC      considerably less efficiently exported from such cells by the multidrug
CC      resistance protein (MDR, or P-glycoprotein). Despite the potential of
CC      epoethiones as anticancer agents, they are problematical to produce on a
CC      large scale. Epoethiones are too complex for industrial scale chemical
CC      synthesis, and Sorangium cellulosum is difficult to ferment, producing
CC      poor yields of epoethiones. The nucleic acids of the invention may be
CC      used for the recombinant production of epoethiones in a heterologous host
CC      that is more amenable to fermentation.
XX
XX      Sequence 68750 BP; 9596 A; 22458 C; 25537 G; 11159 T; 0 other;
XX
XX      Query Match 100.0%; Score 101; DB 21; Length 68750;
XX      Best Local Similarity 100.0%; Pred. No. 1.4e-21;
XX      Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX      QY 1 CGCGGAGTGGCCGTTACAGGGCGTGATCCATGACACCGGAGCGCTCGATGATGCTGC 60
XX      |||||||
DB      36162 CGCGGAGTGGCCGTTACAGGGCGTGATCCATGACACCGGAGCGCTCGATGATGCTGC 36221
XX
XX      QY 61 TTGATGAGCAGACACCGACCGGCTTCTCGGGGTGCTGCA 101
XX      |||||||
DB      36222 TTGATGAGCAGACCGACCGGCTTCTCGGGGTGCTGCA 36262
XX
XX      RESULT 2
XX      AAA29349
XX      ID AAA29349 standard; DNA; 71989 BP.
XX
XX      AC AAA29349;
XX
XX      DT 12-SEP-2000 (first entry)
XX
XX      DE Sorangium cellulosum epoethione polyketide synthase operon genomic DNA.
XX
XX      KW Epoethione; polyketide synthase; epOA; epOB; epOC; epOD; epOE; epOF;
XX      epOL; epOK; P450 epoxidase; ORFA; ORPB; promoter; enhancer; anti-fungal;
XX      tubulin polymerization assay; anti-tumour; cytostatic; ds.
XX
XX      OS Sorangium cellulosum.
XX
XX      FH Key Location/Qualifiers
```

```
FT CDS 3..992 /*tag= a
FT /label= ORF_A
FT /product= transposase
FT /note= "not part of the PKS"
FT CDS 989..1501 /*tag= b
FT /label= ORF_B
FT /product= transposase
FT /note= "not part of the PKS"
FT CDS 1998..6263 /*tag= c
FT /label= epaA_gene
FT /note= "encodes the loading domain"
FT misc_RNA 2031..3548 /*tag= d
FT /note= "encodes ketide synthase (KS-Y) of the loading
FT domain"
FT misc_RNA 3621..4661 /*tag= e
FT /note= "encodes acyl transferase (AT) of the loading
FT domain"
FT misc_RNA 4917..5810 /*tag= f
FT /note= "encodes enoyl reductase (ER) of the loading
FT domain, potentially involved in formation of the
FT thiazole moiety"
FT misc_RNA 5856..6155
FT /*tag= g
FT /note= "encodes acyl carrier protein (ACP) of the loading
FT domain"
FT CDS 6260..10493 /*tag= h
FT /label= epOB_gene
FT /note= "encodes module 1, the NRPS module"
FT misc_RNA 2031..3548 /*tag= i
FT /note= "encodes condensation domain C2 of the NRPS
FT module"
FT misc_RNA 2031..3548
FT /*tag= j
FT /note= "encodes condensation domain C2 of the NRPS
FT module"
FT misc_RNA 6861..6887 /*tag= k
FT /note= "encodes heterocyclization signature sequence"
FT misc_RNA 6861..6887 /*tag= l
FT /note= "encodes condensation domain C4 of the NRPS
FT module"
FT misc_RNA 7358..7366 /*tag= m
FT /note= "encodes condensation domain C7 (partial) of the
FT NRPS module"
FT misc_RNA 7898..7921 /*tag= n
FT /note= "encodes adenylation domain A1 of the NRPS module"
FT misc_RNA 7898..7921 /*tag= o
FT /note= "encodes adenylation domain A1 of the NRPS module"
FT misc_RNA 8261..8308 /*tag= p
FT /note= "encodes adenylation domain A3 of the NRPS module"
FT misc_RNA 8411..8422 /*tag= q
FT /note= "encodes adenylation domain A4 of the NRPS module"
FT misc_RNA 8861..8905 /*tag= r
FT /note= "encodes adenylation domain A6 of the NRPS module"
FT misc_RNA 8966..8983 /*tag= s
FT /note= "encodes adenylation domain A7 of the NRPS module"
FT misc_RNA 9090..9179
FT /*tag= t
FT /note= "encodes adenylation domain A8 of the NRPS module"
FT misc_RNA 9183..9992 /*tag= u
FT /note= "encodes oxidation region for forming thiazole"
FT misc_RNA 10121..10138 /*tag= v
FT /note= "encodes adenylation domain A10 of the NRPS
FT module"
FT misc_RNA 10261..10306 /*tag= w
FT /note= "encodes thiolation domain (PCP) of the NRPS
FT module"
FT CDS 10639..16137 /*tag= x
FT /label= epOC_gene
FT /note= "encodes module 2"
FT misc_RNA 10654..12033 /*tag= y
FT /note= "encodes KS2, the KS domain of module 2"
FT misc_RNA 12250..13287 /*tag= z
FT /note= "encodes AT2, the AT domain of module 2"
FT misc_RNA 13327..13899 /*tag= aa
FT /note= "encodes dehydratase (DH) 2, the DH domain of
FT module 2"
FT misc_RNA 14962..15756 /*tag= ab
FT /note= "encodes ketoreductase (KR) 2, the KR domain of
FT module 2"
FT misc_RNA 15763..16008 /*tag= ac
FT /note= "encodes ACP2, the ACP domain of module 2"
FT CDS 16134..37907 /*tag= ad
FT /label= epOD_gene
FT /note= "encodes modules 3-6"
FT misc_RNA 16425..17606 /*tag= ae
FT /note= "encodes KS3"
FT misc_RNA 17817..18857 /*tag= af
FT /note= "encodes AT3"
FT misc_RNA 19581..20396 /*tag= ag
FT /note= "encodes KR3"
FT misc_RNA 20424..20642 /*tag= ah
FT /note= "encodes ACP3"
FT misc_RNA 20706..22082 /*tag= ai
FT /note= "encodes KS4"
FT misc_RNA 22296..23336 /*tag= aj
FT /note= "encodes AT4"
FT misc_RNA 24069..24647 /*tag= ak
FT /note= "encodes KR4"
FT misc_RNA 24867..25151 /*tag= al
FT /note= "encodes ACP4"
FT misc_RNA 25203..26576 /*tag= am
FT /note= "encodes KS5"
FT misc_RNA 26793..27883 /*tag= an
FT /note= "encodes AT5"
FT misc_RNA 27966..28574 /*tag= ao
FT /note= "encodes DH5"
FT misc_RNA 29433..30287 /*tag= ap
```

| | |
|----|----------------------------------|
| FT | /note= "encodes ER5" |
| FT | 30321..30869 |
| FT | /*tag= aq |
| FT | /note= "encodes KR5" |
| FT | 31077..31373 |
| FT | /*tag= ar |
| FT | /note= "encodes ACP5" |
| FT | 31440..32807 |
| FT | /*tag= as |
| FT | /note= "encodes KS6" |
| FT | 33018..34067 |
| FT | /*tag= at |
| FT | /note= "encodes AT6" |
| FT | 34107..34676 |
| FT | /*tag= au |
| FT | /note= "encodes DH6" |
| FT | 35760..36641 |
| FT | /*tag= av |
| FT | /note= "encodes ER6" |
| FT | 36705..37256 |
| FT | /*tag= aw |
| FT | /note= "encodes KR6" |
| FT | 37470..37769 |
| FT | /*tag= ax |
| FT | /note= "encodes ACP6" |
| FT | 37912..49308 |
| FT | /*tag= ay |
| FT | /label= epoE_gene |
| FT | /note= "encodes modules 7 and 8" |
| FT | 38014..39375 |
| FT | /*tag= az |
| FT | /note= "encodes KS7" |
| FT | 39589..40626 |
| FT | /*tag= ba |
| FT | /note= "encodes AT7" |
| FT | 41341..41922 |
| FT | /*tag= bb |
| FT | /note= "encodes KR7" |
| FT | 42181..42423 |
| FT | misc_RNA |
| FT | 30321..30869 |
| FT | /*tag= aq |
| FT | /note= "encodes KR5" |
| FT | 31077..31373 |
| FT | /*tag= ar |
| FT | /note= "encodes ACP5" |
| FT | 31440..32807 |
| FT | /*tag= as |
| FT | /note= "encodes KS6" |
| FT | 33018..34067 |
| FT | /*tag= at |
| FT | /note= "encodes AT6" |
| FT | 34107..34676 |
| FT | /*tag= au |
| FT | /note= "encodes DH6" |
| FT | 35760..36641 |
| FT | /*tag= av |
| FT | /note= "encodes ER6" |
| FT | 36705..37256 |
| FT | /*tag= aw |
| FT | /note= "encodes KR6" |
| FT | 37470..37769 |
| FT | /*tag= ax |
| FT | /note= "encodes ACP6" |
| FT | 37912..49308 |
| FT | /*tag= ay |
| FT | /label= epoE_gene |
| FT | /note= "encodes modules 7 and 8" |
| FT | 38014..39375 |
| FT | /*tag= az |
| FT | /note= "encodes KS7" |
| FT | 39589..40626 |
| FT | /*tag= ba |
| FT | /note= "encodes AT7" |
| FT | 41341..41922 |
| FT | /*tag= bb |
| FT | /note= "encodes KR7" |
| FT | 42181..42423 |
| FT | misc_RNA |

| | | | | |
|-----------------------|---|--|-----------|---------------|
| Query Match | 100.0% | Score 101; | DB 21; | Length 71989; |
| Best Local Similarity | 100.0% | Pred. No. 1,4e-21; | | |
| Matches 101; | Conservative 0; | Mismatches 0; | Indels 0; | Gaps 0; |
| OY | 1 | CGCGGAGTGGCCGTTACAGGGCGTGATCCATGCACGCCGAGCCGCTCGATGATGTGTGC | 60 | |
| | | | | |
| Db | 30550 | CGCGCGAGTGGCCGTTACAGGGCGTGATCCATGCACGCCGAGCCGCTCGATGATGTGTGC | 30609 | |
| OY | 61 | TTGATGAGCAGCACCGCAGCCGCTTCTCGCGGCTCTGSCA | 101 | |
| | | | | |
| Db | 30610 | TTGATGAGCAGCACCGCAGCCGCTTCTCGCGGCTCTGSCA | 30650 | |
| RESULT 3 | | | | |
| AA517367 | | | | |
| ID | AA517367 | standard; DNA; 33529 BP. | | |
| XX | | | | |
| XX | AA517367; | | | |
| XX | | | | |
| DT | 12-MAR-2002 | (first entry) | | |
| XX | | | | |
| DE | DNA sequence of <i>S. cellulosum</i> polyketide synthase cosmid, pKOS28-26. | | | |
| XX | | | | |
| KW | Polyketide synthase; PKS; catalytic domain; ketosynthase domain; | | | |
| KW | acyl transferase domain; dehydratase domain; ketoreductase domain; | | | |
| KW | acyl carrier protein domain; pKOS28-26; ds. | | | |
| XX | | | | |
| XX | | | | |
| OS | <i>Sorangium cellulosum</i> . | | | |
| XX | | | | |
| PN | US6280999-B1. | | | |
| XX | | | | |
| PD | 28-AUG-2001. | | | |

PE 31-AUG-1998; 98US-0144085.
 XX
 PR 22-JAN-1998; 98US-0010809.
 XX
 PA (KOSA-) KOSAN BIOSCIENCE.
 XX
 PI Gustafsson C, Betlach MC, Ashley G, Julien B, Ziermann R;
 DR WPI; 2001-606536/69.
 XX
 XX
 PR Novel purified, isolated DNA molecule from *Sorangium cellulosum* having
 PT polyketide open reading frame encoding modules with one or more domains
 PR such as ketosynthase, acyl transferase and acyl carrier protein domains
 XX
 XX
 PS Claim 4; Fig 1; 72pp; English.

The present invention relates to the isolation of novel *Sorangium* cellulose polysaccharide synthases (PKS), and the polynucleotide sequences encoding them. The polysaccharide synthases include catalytic domains such as ketosynthase domain, acyl transferase domain, dehydratase domain, ketoreductase domain, and acyl carrier protein domain. A host cell comprising a PKS ORF (open reading frame) which encodes one or more PKS domains is useful for producing polysaccharide synthases from which polysaccharides can be produced. The host cells are useful for constructing a library, where each individual colony of the library represents a colony with the ability to produce a particular PKS synthase and ultimately a particular polysaccharide. The polysaccharides produced by these colonies can be used collectively in a panel to represent a library or may be assessed individually for activity. Colonies in the library are also induced to produce the relevant synthases and thus to produce the relevant polysaccharides to obtain a library of candidate polysaccharides which can be screened for binding to desired targets such as receptors, signalling proteins, etc. The present sequence represents the DNA sequence of cosmid PKOS28-26 which encodes one or more domains of *S. cellulose* PKS.

Note: The present sequence is said to encode the functional domains of *S. cellulose* PKS which correspond to domains or domain subsets of the amino acid sequences of ORF1 (AA010700) and ORF2 (AA010701).

| Seq | Sequence | 33529 BP | 4489 A | 9518 C | 14470 G | 5046 T | 6 other |
|-----------------------|---|--|------------|--------|---------|--------|---------|
| Query Match | 51.1% | Score | 51.6 | DB | 23 | Length | 33529 |
| Best Local Similarity | 70.4% | Pred. No. | 2.3e-06 | | | | |
| Matches 69 | Conservative | 0 | Mismatches | 29 | Indels | 0 | Gaps |
| QY | 3 | GGCGAGTGGCCCTTACAGAGCGCGTGTATCATGTGACGCGAGCGCCTCGATGATGTGTGCTT | 62 | | | | |
| Db | 14155 | GGCGGGCGCCGCTGTGAGCGCGGTGCTGTGACACTGGCCGGCGGCGCCTGTGACGACGCGGTGCTC | 14214 | | | | |
| | | | | | | | |
| QY | 63 | GATGAGCAGACACGACCGACCGCTTCTTCGCGGGTGTCTGGC | 100 | | | | |
| Db | 14215 | GCCGGCCAGACGCGCGACGCGCTCTCGCGGGTGTGGC | 14252 | | | | |
| | | | | | | | |
| RESULT 4 | | | | | | | |
| AAA92302 | | | | | | | |
| ID | AAA92302 | standard | DNA | 31422 | BP | | |
| XX | AAA92302 | | | | | | |
| AC | | | | | | | |
| XX | | | | | | | |
| DT | 10-JAN-2001 | (first entry) | | | | | |
| XX | | | | | | | |
| DE | S. avermectilis avermectin aglycon synthase DNA avealiI SEQ ID NO:2. | | | | | | |
| XX | | | | | | | |
| KM | Streptomyces avermectilis; avermectin aglycon synthase; biosynthesis; | | | | | | |
| KW | multifunctional enzyme; polyketide; avermectin; veterinary drug; | | | | | | |
| KW | agrochemical; ds. | | | | | | |
| XX | | | | | | | |
| OS | Streptomyces avermectilis. | | | | | | |
| XX | | | | | | | |
| Key | | Location/Qualifiers | | | | | |
| FT | CDS | 1..14646 | | | | | |

```
FT      /*tag= a
FT      /note= "avermectin aglycon synthase protein"
FT      CDS      14824..31422
FT      /*tag= b
FT      /note= "avermectin aglycon synthase protein"
XX      MO200050605-A1.
XX      PD      31-AUG-2000.
XX      PE      23-FEB-2000; 2000WO-JP01041.
XX      PR      24-FEB-1999; 99JP-0046961.
XX      PA      (KITA ) KITASATO INST.
XX      PI      Omura S, Ikeda H;
XX      DR      WPI; 2000-565458/52.
XX      PT      P-PSDB; AAB23751, AAB23752.
XX      PS      Claim 2; Page 134-203; 314pp; Japanese.
XX      CC      The present sequence represents DNA which encodes avermectin aglycon
XX      CC      synthase proteins. Also described are: (1) polypeptides encoded by all
XX      CC      or part of the DNA; (2) expression vectors containing the DNA; (3) host
XX      CC      cells transformed by the vectors; (4) preparation of the polypeptides
XX      CC      by culture of the transformants; (5) preparation of avermectin aglycon
XX      CC      or its derivatives by culture of transformed avermectin-producing
XX      CC      microorganisms; and (6) oligonucleotides of 5-60 bases in length
XX      CC      containing sense or antisense sequences from the avermectin aglycon
XX      CC      synthase DNA. The enzymes are useful for the production of modified
XX      CC      forms of avermectin and of the intermediates in its biosynthesis, for
XX      CC      use as drugs, veterinary drugs and agrochemicals.
XX      SQ      Sequence 31422 BP; 4136 A; 10238 C; 11677 G; 5371 T; 0 other;
SQ      Query Match      46.3%; Score 46.8; DB 21; Length 31422;
SQ      Best Local Similarity 67.3%; Pred. No. 7.1e-05;
SQ      Matches 66; Conservative 0; Mismatches 32; Indels 0; Gaps 0;
OY      1 CGGCGAGTGGCCGTTACAGGCGGTGATCATGCAGCCGAGCGCTCGATGATGCTGCG 60
OY      ||||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      3863 CGGCGAGTGGCCGCTGTGGTGTGTCATGCCGCGAGTGTGTGATGACGGGCTGT 3922
OY      61 TTGATGAGCAGACACGACGCGCTTCTCGCGGGTCTG 98
OY      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      3923 TGGAGGGTGTGTCGTCGAGCGTGTACAGGGGGTCTG 3960

RESULT 5
AAH79278
ID      AAH79278 standard; DNA; 31422 BP.
XX      AC      AAH79278;
XX      DT      04-DEC-2001 (first entry)
XX      DE      Streptomyces avermitilis coding sequences SEQ ID NO: 2.
XX      KW      Avermectin aglycone synthase; AAs; avermectin derivative;
XX      KW      drug production; veterinary drug; pesticide; ds.
XX      OS      Streptomyces avermitilis.
XX      FH      Key      Location/Qualifiers
XX      FT      CDS      1..14646
XX      FT      /*tag= a
XX      FT      /product= "AAG65266"
```

```
FT      CDS      14824..31422
FT      /*tag= b
FT      /product= "AAG65267"
XX      MO200162939-A1.
XX      PD      30-AUG-2001.
XX      PE      23-FEB-2001; 2001WO-JP01381.
XX      PR      24-FEB-2000; 2000JP-0047405.
XX      PA      (KYOW ) KYOWA HAKKO KOGYO KK.
XX      PA      (KITA ) KITASATO INST.
XX      PI      Endo H, Yamaguchi H, Kanda Y, Hashimoto S, Omura S, Ikeda H;
XX      DR      WPI; 2001-582053/65.
XX      PT      P-PSDB; AAG65266, AAG65267.
XX      PS      New modified avermectin aglycone synthase derived from Streptomyces
XX      PS      avermectilis used in production of 22,23-dihydroavermectin B1a used in
XX      PS      drugs and pesticides -
XX      PS      Disclosure; Page 103-149; 257pp; Japanese.
XX      CC      The present invention relates to the production of modified derivatives
XX      CC      of avermectin aglycone synthase (AAs) derived from Streptomyces
XX      CC      avermectilis. The activity of an acyl carrier protein (ACP),
XX      CC      beta-ketothiol carrier protein synthase (KS), acyltransferase (AT),
XX      CC      beta-ketocacyl carrier protein reductase (KR), dehydratase (DH), enoyl
XX      CC      reductase (ER) and/or thioesterase (TE) domain may be reduced or
XX      CC      suppressed. The process can be used in the production of drugs, veterinary
XX      CC      drugs and pesticides. The present sequence is a fragment of the S.
XX      CC      avermectilis genome.
XX      SQ      Sequence 31422 BP; 4136 A; 10237 C; 11677 G; 5372 T; 0 other;
SQ      Query Match      46.3%; Score 46.8; DB 22; Length 31422;
SQ      Best Local Similarity 67.3%; Pred. No. 7.1e-05;
SQ      Matches 66; Conservative 0; Mismatches 32; Indels 0; Gaps 0;
OY      1 CGGCGAGTGGCCGTTACAGGCGGTGATCATGCAGCCGAGCGCTCGATGATGCTGCG 60
OY      ||||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      3863 CGGCGAGTGGCCGCTGTGGTGTGTCATGCCGCGAGTGTGTGATGACGGGCTGT 3922
OY      61 TTGATGAGCAGACACGACGCGCTTCTCGCGGGTCTG 98
OY      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      3923 TGGAGGGTGTGTCGTCGAGCGTGTACAGGGGGTCTG 3960

RESULT 6
AAV05287
ID      AAV05287 standard; DNA; 49377 BP.
XX      AC      AAV05287;
XX      DT      21-MAY-1998 (first entry)
XX      DE      The soraphen biosynthesis gene cluster from Sorangium cellulosum.
XX      KW      Polyketide synthase; PKS; biosynthesis; soraphen; SOR; SORa; SORb;
XX      KW      SORn; biosynthetic module; beta-ketocacylsynthase; acyltransferase;
XX      KW      ketoreductase; beta-ketone processing domain; cytosstatic agent;
XX      KW      antimicrobial agent; phytopathogenic fungi; transgenic plant;
XX      KW      biological control; ss.
XX      OS      Sorangium cellulosum.
XX      FH      Key      Location/Qualifiers
XX      FT      CDS      383..760
XX      FT      /*tag= a
XX      FT      /product= SORR
```

FT /note= "gene product highly homologous to the
 FT reductase domains of type I PKs such
 FT as erya from Saccharopolyspora erythraea"
 FT CDS 927..19874
 FT /tag= b
 FT /product= Sora
 FT /note= "gene product is highly homologous to
 FT type I PKs that are known to be involved
 FT in the synthesis of polyketide compounds"
 FT misc_feature 942..7115
 FT /tag= c
 FT /note= "module 1 of Sora"
 FT misc_feature 7203..12884
 FT /tag= d
 FT /note= "module 2 of Sora"
 FT misc_feature 13455..19616
 FT /tag= e
 FT /note= "module 3 of Sora"
 FT misc_feature 19871..46318
 FT /tag= f
 FT CDS 19871..46318
 FT /tag= g
 FT /product= Sora
 FT /note= "gene product is highly homologous to
 FT type I PKs genes"
 FT misc_feature 19870..24556
 FT /tag= h
 FT /note= "module 1 of Sora"
 FT misc_feature 24638..30820
 FT /tag= i
 FT /note= "module 2 of Sora"
 FT misc_feature 30881..35446
 FT /tag= j
 FT /note= "module 3 of Sora"
 FT misc_feature 35528..40114
 FT /tag= k
 FT /note= "module 4 of Sora"
 FT misc_feature 40190..46318
 FT /tag= l
 FT /note= "module 5 of Sora"
 FT CDS 46851..47891
 FT /tag= m
 FT /product= Sorm
 FT /note= "gene product is homologous to the
 FT methyltransferase from Streptomyces
 FT hygroscopicus that is involved in
 FT the synthesis of the polyketide rapamycin"
 FT US5716849-A.
 FT PD 10-FEB-1998.
 FT XX 14-DEC-1996; 96US-0764233.
 FT XX 24-AUG-1993; 93WO-US07954.
 FT PR 08-JUN-1994; 94US-0258261.
 FT PR 09-OCT-1996; 96US-0729214.
 FT XX (NOVS) NOVARTIS FINANCE CORP.
 FT PA Beck JT, Hill DS, Ligon JM, Neff S, Ryals JA, Schupp T;
 FT PI WPI, 1998-158369/14.
 FT DR WPI, 1998-158369/14.
 FT XX DNA encoding Sorangium cellulosum polypeptide(s) - used for, e.g.
 FT PT biosynthesis of soraphen useful as antimicrobial agent against
 FT PT phytopathogenic fungi
 FT XX Claim 2; Columns 47-90; 64pp; English.
 FT PS The present sequence contains a cluster of genes that encode polyketide
 CC synthases (PKSs) that are involved in the synthesis of soraphens in
 CC Sorangium cellulosum. The proteins encoded by the present sequence are

CC Sora, Sora and Sorm. Sora and Sora contain biosynthetic modules
 CC which contain a beta-ketoacylsynthase, an acyltransferase, a
 CC ketoreductase and an acyl carrier protein domain, as well as beta-ketone
 CC processing domains. S. cellulosum soraphens are useful as a cytostatic
 CC and antimicrobial agent active against phytopathogenic fungi.
 CC Soraphen-producing transgenic plants or biological control agents can
 CC also be produced, which may reduce crop losses and nutritional
 CC deprivation for local populations in many parts of the world.
 CC XX
 SO Sequence 49377 BP: 7247 A; 19522 C; 14477 G; 8131 T; 0 other;
 Query Match 43.6%; Score 44; DB 19; Length 49377;
 Best Local Similarity 65.0%; Pred. No. 0.00056;
 Matches 65; Conservative 0; Mismatches 35; Indels 0; Gaps 0;
 OY 1 CGCGGAGTGGCCGCTTACGAGCGGCTGATCCATCCAGCGGCGCTGCATGTCGTCG 60
 DB 6277 CGTCGGACGTCCCGCCGTCACGCGCGTCTTCACAGCGGACGAGCTCGCAGATGGCTGA 6336
 OY 61 TTGATGACGACGACACCGACCGCTTCTCGCGGTCGTCG 100
 DB 6337 TCGCGCACATGACGCGCGGACGCGATGACGCGGTCCTTGC 6376
 RESULT 7
 AAT80413
 ID AAT80413 standard; DNA; 43280 BP.
 AC AAT80413;
 DT 27-FEB-1998 (first entry)
 DE Tyactone synthase gene cluster.
 KW Tyactone synthase gene cluster; tylG gene; multifunctional protein;
 KW polyketide; tyactone synthesis; antibiotic; tylosin; ss.
 OS Streptomyces fradiae.
 FH Key Location/Qualifiers
 FT CDS 816..14243
 FT /tag= a
 FT /transl_except= (pos: 816..818, aa: Met)
 FT /note= "ORF1 encodes protein shown in AAW22601"
 FT CDS 14351..19945
 FT /tag= b
 FT /transl_except= (pos: 14351..14353, aa: Met)
 FT /note= "ORF2 encodes protein shown in AAW22602"
 FT CDS 20010..31199
 FT /tag= c
 FT /transl_except= (pos: 20010..20012, aa: Met)
 FT /note= "ORF3 encodes protein shown in AAW22603"
 FT CDS 31232..36067
 FT /tag= d
 FT /note= "ORF4 encodes protein shown in AAW22604"
 FT CDS 36249..41774
 FT /tag= e
 FT /note= "ORF5 encodes protein shown in AAW22605"
 FT EP791655-A2.
 FT PD 27-AUG-1997.
 FT XX 19-FEB-1997; 97EP-0301056.
 FT PR 22-FEB-1996; 96US-0012078.
 FT XX (ELIL) LILLY & CO ELI.
 FT PI Dehoff BS, Kuhstoss SA, Rosteck PR, Sutton KL;
 FT XX WPI, 1997-418046/39.
 DR P-PSDB; AAW22601-W22605.

XX DNA encoding Streptomyces fradiae tylactone synthase domain - for
PT production of tylosin-related polyketide compounds
XX
XX Claim 2; Pages 8-66; 220pp; English.
XX
CC This sequence represents the tylactone synthase gene cluster of the
CC invention. This sequence is also referred to as the tylG gene, and was
CC isolated from Streptomyces fradiae. This sequence encodes multifunctional
CC proteins which direct the synthesis of the polyketide tylactone. Isolated
CC from Streptomyces fradiae. Tylactone is the basic building block of the
CC antibiotic tylosin. The DNA sequence can be modified so as to alter the
CC type of carboxylic acids incorporated, the number of carboxylic acids
CC incorporated and/or the post-condensation reactions performed, thereby
CC resulting in novel tylosin-related polyketides.
XX
SQ Sequence 43280 BP; 5269 A; 15187 C; 16651 G; 6173 T; 0 other;
Query Match 40.8%; Score 41.2; DB 18; Length 43280;
Best Local Similarity 67.4%; Pred. No. 0.0041;
Matches 58; Conservative 0; Mismatches 28; Indels 0; Gaps 0;
QY 1 CGGCGAGTGGCCGTTACAGGCGGTGATCATGACGCGGAGCGCTGATGATGGTGTGC 60
DB 13093 CGGCCGACCGCGCGCTGACCGAGATGATCAGCGCGCGGTGCTCGACGCGCACAC 13152
QY 61 TTGATGACGACAGCACCGACCGCTTC 86
DB 13153 TCGACGCGCTGACCGCGGAGACGATC 13178

RESULT 8
AADI7186
ID AADI7186 standard; DNA; 125401 BP.
XX
XX AADI7186;
AC
XX 29-NOV-2001 (first entry)
DT
XX
DE Streptomyces noursei nystatin PKS gene cluster DNA.
XX
XX Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KW antifungal; antibiotic; ds.
XX
XX Streptomyces noursei.
OS

XX FH Location/Qualifiers
FT Key 6337..34771
FT CDS
FT /*tag= a
FT /product= "NysI complete protein"
FT 34792..51099
FT /*tag= b
FT /product= "NysJ protein"
FT 51155..57355
FT /*tag= c
FT /product= "NysK protein"
FT 57503..58687
FT /*tag= d
FT /product= "NysL protein"
FT complement (58786..58980)
FT /*tag= e
FT /product= "NysM protein"
FT /note= "CDS does not include start codon"
FT complement (59045..60241)
FT /*tag= f
FT /product= "NysN protein"
FT /note= "CDS does not include start codon"
FT complement (60238..61296)
FT /*tag= g
FT /product= "NysD2 complete protein"
FT 120628..121308
FT /*tag= h
FT /product= "NysR4 (long) protein"

XX
PN WO200159126-A2.
XX
XX 16-AUG-2001.
PD
XX
XX 08-FEB-2001; 2001WO-CB00509.
PF
XX
XX 08-FEB-2000; 2000GB-0002840.
PR 10-APR-2000; 2000GB-0008786.
PR 14-APR-2000; 2000GB-0009387.
XX
PA (UNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIG.
PA (SMTF) SINTEF SPIFTELSEN IND TEK FORSK.
PA (ALPH-) ALPHARMA AS.
PA (SINV-) SINVENT AS.
PA (DZIE/) DZIELEWSKA H.
PA (ZOTC/) ZOTCHEV S B.
PA (SEKU/) SEKUROVA O N.
PA (FJAE/) FJAEVYIK E.
PA (BRAU/) BRAUTASET T.
PA (STRO/) STROM A R.
XX
PI Zotchev SB, Sekurova ON, Fjaevyik E, Brautaset T, Strom AR;
PI Valla S, Ellingsen TE, Sletta H, Gulliksen O;
XX
DR WPI: 2001-557614/62.
DR P-PSDB: AAE10143, AAE10144, AAE10145, AAE10146, AAE10147, AAE10148,
DR AAE10149, AAE10150.
XX
PT New nystatin polyketide synthase polynucleotides and polypeptides,
PT useful as antibiotics and antifungals -
XX
PS Claim 1; Page 188-254; 266pp; English.
XX
CC The present invention relates to the cloning and sequencing of the gene
CC cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC involved in the biosynthesis of the macrolide antibiotic nystatin.
CC The nystatin PKS is useful as antifungal antibiotics. The present
CC sequence is a Streptomyces noursei nystatin PKS gene cluster DNA.
XX
SQ Sequence 125401 BP; 15664 A; 49692 C; 42871 G; 17174 T; 0 other;
Query Match 40.8%; Score 41.2; DB 22; Length 125401;
Best Local Similarity 64.9%; Pred. No. 0.0049;
Matches 61; Conservative 0; Mismatches 33; Indels 0; Gaps 0;
QY 1 CGGCGAGTGGCCGTTACAGGCGGTGATCATGACGCGGAGCGCTGATGATGGTGTGC 60
DB 44786 CGCGGAGTGGCCCGCTGACCGCGGTGATCATGACCGCGGAGTGTCTGACGACGCGGTGC 44845
QY 61 TTGATGACGACAGCACCGACCGCTTCTGCGGGGT 94
DB 44846 TCACCGGCTTCACCCCGACCGGCTTCTGCGACCGT 44879

RESULT 9
AAF90038
ID AAF90038 standard; DNA; 5088 BP.
XX
XX AAF90038;
AC
XX
XX 06-AUG-2001 (first entry)
DT
XX
DE Nucleotide sequence of a type I polyketide synthase.
XX
XX Metabolic pathway operon; polyketide; polyketide antibiotic;
KW type I polyketide synthase; ss.
XX
OS Unidentified.
XX
XX FH Location/Qualifiers
FT Key 1..5088
FT CDS
FT /*tag= a

```

FT      /product- "type I polyketide synthase"
FT      /trans_except- "(1..3, aa: Met)"
XX
XX
PN      MO200140497-A2.
XX
XX      07-JUN-2001.
XX
XX      27-NOV-2000; 2000WO-FR03311.
XX
XX      29-NOV-1999; 99FR-0015032.
XX      07-JUN-2000; 2000US-0209800.
XX
XX      (AVET ) AVENTIS PHARMA SA.
XX
XX      Jeanin P, Pernodet J, Guerineau M, Simonet P, Courtois S;
XX      Cappellano C, Franco F, Raynal A, Ball M, Sezonov G, Tuphille K;
XX      Frostegard A;
XX
XX      WPI; 2001-374849/39.
XX      P-PSDB; AAB83975.
XX
XX      Collection of nucleic acids from environmental samples, useful for
XX      identifying e.g. genes encoding polyketide synthases and derived
XX      antibiotics
XX
XX      Claim 35; Page 309-311; 356pp; French.
XX
XX      The specification describes a method for the preparation of a collection
XX      of nucleic acids from organisms in a soil sample. The method comprises
XX      milling a dried sample to produce microparticles; suspending these in
XX      liquid buffer; extraction of nucleic acids from the microparticle;
XX      passing nucleic acid-containing solution through a molecular sieve;
XX      passing nucleic acid-enriched fractions through an anion exchange
XX      chromatography material; and recovering fractions containing purified
XX      nucleic acids. The nucleic acids are sources for sequences that encode
XX      either operators involved in a metabolic pathway (specifically polyketide
XX      synthetase) or polypeptides, particularly for production of therapeutic
XX      or agricultural compounds, especially polyketide antibiotics. AAF90034-39
XX      represent open reading frames (ORFs) of the coding strand of cosmid
XX      a26g1, and encode type I polyketide synthases.
XX
XX      Sequence 5088 BP; 838 A; 1622 C; 1679 G; 949 T; 0 other;
XX
XX      Query Match      40.0%; Score 40.4; DB 22; Length 5088;
XX      Best Local Similarity 65.6%; Pred. No. 0.0049;
XX      Matches 59; Conservative 0; Mismatches 31; Indels 0; Gaps 0;
XX
XX      OY      11 GCGGTACAGGCGGTATCATCCAGCGGCTCGATGATGTGCTTATGAGCA 70
XX      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
XX      DB      4239 GCGGCTGCGTGGAGTGTGTCATGCCAGCGGTCTGATGAGGTAAGTACAGACA 4298
XX
XX      OY      71 GACCAACGACCGGCTCTCGCGGGTGTGGC 100
XX      ||| | | | | | | | | | | | | | | | | | | | | | | | | |
XX      DB      4299 GACGTGGCTCTTTCGAGAGAGGTCTGGC 4328
XX
XX      RESULT 10
XX      AAF24892
XX      ID      AAF24892 standard; DNA; 20394 BP.
XX
XX      AC      AAF24892;
XX
XX      DT      20-APR-2001 (first entry)
XX
XX      DE      Pimaricin biosynthesis associated polyketide synthase gene.
XX
XX      KM      Polyketide synthase; oxidative modification; metabolite; antibiotic;
XX      anticancer; pimaricin; ss.
XX
XX      OS      Streptomyces natalensis.
XX
XX      Key      Location/Qualifiers
XX      CDS      1..20394
XX

```

```

FT      /*tag- a
FT      /product- "polyketide synthase"
XX
XX
PN      MO200077222-A1.
XX
XX      21-DEC-2000.
XX
XX      14-JUN-2000; 2000WO-EP06227.
XX
XX      14-JUN-1999; 99EP-0201893.
XX
XX      (STAM ) DSM NV.
XX
XX      Martin JF, Aparicio JF, Colina AJ;
XX
XX      WPI; 2001-080693/09.
XX      P-PSDB; AAB31558.
XX
XX      New polynucleotides encoding enzymes involved in the biosynthesis of
XX      pimaricin, useful for modifying the biosynthesis of pimaricin and in
XX      the synthesis of new compounds
XX
XX      Disclosure; Page 53-80; 116pp; English.
XX
XX      The present sequence encodes a polyketide synthase which is associated
XX      with the biosynthesis of pimaricin. The polyketide synthase polypeptide
XX      is useful for the oxidative modification of a methyl group of a suitable
XX      compound, e.g. a bioactive compound including a secondary metabolite,
XX      antibiotics and anticancer agents. Recombinant cells comprising the
XX      gene are useful for the production of pimaricin. The polyketide synthase
XX      polynucleotide may be over expressed in Streptomyces, leading to an
XX      increase in the biosynthesis of pimaricin, as a source of primers for
XX      amplification reaction and as probes.
XX
XX      Sequence 20394 BP; 2611 A; 7486 C; 7397 G; 2900 T; 0 other;
XX
XX      Query Match      40.0%; Score 40.4; DB 22; Length 20394;
XX      Best Local Similarity 63.3%; Pred. No. 0.0063;
XX      Matches 62; Conservative 0; Mismatches 36; Indels 0; Gaps 0;
XX
XX      OY      1 CGCGGAGATGCGCGTTACAGGCGGTATCATCCAGCGGCTCGATGATGTGTCG 60
XX      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
XX      DB      3797 CGCGGACCTGCGGTACCGGCGGTCTCCACGCCGCGGTACGAGAGCGGCTAC 3856
XX
XX      OY      61 TTGATGACAGACACCGACCGCTCTCGCGGGTGTG 98
XX      ||||| | | | | | | | | | | | | | | | | | | | | | | | |
XX      DB      3857 TGGACGCGCTACCCCGAGCGCTTCGAGACCGTACTG 3894
XX
XX      RESULT 11
XX      AAF90033
XX      ID      AAF90033 standard; DNA; 34071 BP.
XX
XX      AC      AAF90033;
XX
XX      DT      06-AUG-2001 (first entry)
XX
XX      DE      Nucleotide sequence of cosmid a26g1 (coding strand).
XX
XX      KM      Metabolic pathway operon; polyketide; polyketide antibiotic; ss.
XX
XX      OS      Synthetic.
XX
XX      PN      MO200140497-A2.
XX
XX      PD      07-JUN-2001.
XX
XX      PF      27-NOV-2000; 2000WO-FR03311.
XX
XX      PR      29-NOV-1999; 99FR-0015032.
XX      07-JUN-2000; 2000US-0209800.
XX
XX      (AVET ) AVENTIS PHARMA SA.
XX

```


XX Jeanin P, Peronnet J, Guerin M, Simonet P, Courtois S;
PI Cappellano C, Francou F, Raynal A, Ball M, Sezonov G, Tuphile K;
PI Frostegard A;
XX WPI: 2001-374849/39.
XX
XX Collection of nucleic acids from environmental samples, useful for
PT identifying e.g. genes encoding polyketide synthases and derived
PT antibiotics
XX
XX Example 14; Page 289-300; 356pp; French.
XX
XX The specification describes a method for the preparation of a collection
CC of nucleic acids from organisms in a soil sample. The method comprises
CC milling a dried sample to produce microparticles; suspending these in
CC liquid buffer; extraction of nucleic acids from the microparticles;
CC passing nucleic acid-containing solution through a molecular sieve;
CC passing nucleic acid-enriched fractions through an anion exchange
CC chromatography material; and recovering fractions containing purified
CC nucleic acids. The nucleic acids are sources for sequences that encode
CC either operons involved in a metabolic pathway (specifically polyketide
CC synthesis) or polypeptides, particularly for production of therapeutic
CC or agricultural compounds, especially polyketide antibiotics. The present
CC sequence represents cosmid a26g1 coding strand, which encodes different
CC polyketide synthases.
XX
XX
SQ Sequence 34071 BP; 5791 A; 10858 C; 11089 G; 6333 T; 0 other;
Query Match 40.0%; Score 40.4; DB 22; Length 34071;
Best Local Similarity 65.6%; Pred. No. 0.0069;
Matches 59; Conservative 0; Mismatches 31; Indels 0; Gaps 0;
OY 11 GCCGTACAGGCGCGATGCATGCAGCGGCGCTCGATGATGCTTGATGACA 70
DB 28920 GCCGCTGGGTGAGTGCATGCGCGGCGCTCGATGCGGCGTATGACAACA 28979
OY 71 GACCACGACGCGCTTCGCGGGGTCTGGC 100
DB 28980 GACGTGGGCTCTTTCGAGAGGTCTGGC 29009

RESULT 12
AAF90032/C
ID AAF90032 standard; DNA; 42717 BP.
XX
XX AAF90032;
XX
XX 06-AUG-2001 (first entry)
XX
XX Nucleotide sequence of cosmid a26g1 (non-coding strand).
XX
XX Metabolic pathway operon; polyketide; polyketide antibiotic; ss.
XX
XX Synthetic.
XX
XX WO200140497-A2.
XX
XX 07-JUN-2001.
XX
XX 27-NOV-2000; 2000WO-FR0311.
XX
XX 29-NOV-1999; 99FR-0015032.
XX
XX 07-JUN-2000; 2000US-0209800.
XX
XX (AVENTIS PHARMA SA.
XX
XX Jeanin P, Peronnet J, Guerin M, Simonet P, Courtois S;
PI Cappellano C, Francou F, Raynal A, Ball M, Sezonov G, Tuphile K;
PI Frostegard A;
XX
XX WPI: 2001-374849/39.
XX

PT Collection of nucleic acids from environmental samples, useful for
PT identifying e.g. genes encoding polyketide synthases and derived
PT antibiotics
XX
XX Example 14; Page 274-288; 356pp; French.
XX
XX The specification describes a method for the preparation of a collection
CC of nucleic acids from organisms in a soil sample. The method comprises
CC milling a dried sample to produce microparticles; suspending these in
CC liquid buffer; extraction of nucleic acids from the microparticles;
CC passing nucleic acid-containing solution through a molecular sieve;
CC passing nucleic acid-enriched fractions through an anion exchange
CC chromatography material; and recovering fractions containing purified
CC nucleic acids. The nucleic acids are sources for sequences that encode
CC either operons involved in a metabolic pathway (specifically polyketide
CC synthesis) or polypeptides, particularly for production of therapeutic
CC or agricultural compounds, especially polyketide antibiotics. The present
CC sequence represents cosmid a26g1 (non-coding strand). The sense strand
CC encodes different polyketide synthases.
XX
XX
SQ Sequence 42717 BP; 8230 A; 13520 C; 13184 G; 7782 T; 1 other;
Query Match 40.0%; Score 40.4; DB 22; Length 42717;
Best Local Similarity 65.6%; Pred. No. 0.0072;
Matches 59; Conservative 0; Mismatches 31; Indels 0; Gaps 0;
OY 11 GCCGTACAGGCGCGATGCATGCAGCGGCGCTCGATGATGCTTGATGACA 70
DB 5200 GCCGCTGGGTGAGTGCATGCGCGGCGCTCGATGCGGCGTATGACAACA 5141
OY 71 GACCACGACGCGCTTCGCGGGGTCTGGC 100
DB 5140 GACGTGGGCTCTTTCGAGAGGTCTGGC 5111

RESULT 13
AAH52062
ID AAH52062 standard; DNA; 4851 BP.
XX
XX AAH52062;
XX
XX 04-SEP-2001 (first entry)
XX
XX Mycobacterium tuberculosis potential drug target gene SEQ ID 116.
XX
XX Drug target; growth; organism viability; characterisation; ds.
XX
XX Mycobacterium tuberculosis.
XX
XX WO200135317-A1.
XX
XX 17-MAY-2001.
XX
XX 13-NOV-2000; 2000WO-US31152.
XX
XX 12-NOV-1999; 99US-0165086.
XX
XX 12-NOV-1999; 99US-0165124.
XX
XX 01-FEB-2000; 2000US-0179531.
XX
XX (REGC) UNIV CALIFORNIA.
XX
XX Eisenberg D, Rotstein SH, Marcotte EM;
PI P-PSDB; AAG81211.
XX
XX WPI: 2001-329193/34.
XX
XX P-PSDB; AAG81211.
XX
XX Identifying nucleotide or polypeptide sequence for use as drug target,
PT involves providing algorithm that analyses a functional relationship
PT between nucleotide or polypeptide sequences, and comparing the
PT sequences
XX
XX Disclosure; Page 131-133; 207pp; English.
XX

CC This invention relates to a method for identifying a nucleotide or
CC polypeptide sequence that may be a drug target, or essential for growth
CC or viability of an organism. Polynucleotide sequences AAH51947 - AAH52092
CC represent DNA encoding proteins AAG81096 - AAG81241. Mycobacterium
CC tuberculosis proteins which are potential drug targets. The DNA and
CC protein sequences are used to illustrate the method of the invention. The
CC method involves providing an unknown nucleotide or polypeptide sequences,
CC and comparing it to a number of sequences along with at least one
CC algorithm capable of analyzing a functional relationship between
CC nucleotide and polypeptide sequences. The method is useful for
CC characterizing the function of nucleic acids and polypeptides that may be
CC useful as a target for a drug or essential for the growth or viability of
CC an organism.
CC XX

SQ Sequence 4851 BP; 611 A; 1403 C; 1914 G; 923 T; 0 other;

Query Match 39.0%; Score 39.4; DB 22; Length 4851;
Best Local Similarity 62.9%; Pred. No. 0.01;
Matches 61; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 2 GGGGAGTGGCGCTTACAGGGCGTATCGATCGAGCGGCTCGATGCTGCT 61
DB 3915 GGAGTACCCCGCGGTGGGGGATTCATGCCCGCGCTGCTCATACGACGTGAT 3974
QY 62 TGATGACGACGACCGACCGCTTCTCGCGGTGCTG 98
DB 3975 CACCTCGTGCACACGCGACCGCATCGATCGGTG 4011

RESULT 14
AA168063
ID AA168063 standard; DNA; 552 BP.
XX
XX AA168063;
XX
DT 13-MAR-2002 (first entry)
XX
XX Epothilone PKS KR domain nucleotide sequence.
DE
XX
XX Cystobacterinease; recombinant; polyketide synthase; PKS; polyketide;
KW epothilone; p1A gene; cytosolic; antiproliferative; antiarthritic;
KW antitartaric; antitartaric; antitartaric; neuroprotective; vasotropic; ds.
XX
XX Synthetic.
OS
FH Key Location/Qualifiers
FT CDS 1..552
FT /*tag= a
FT /note= "KR domain of extender module 6"

PN WO200183800-A2.
XX
XX 08-NOV-2001.
PD
XX
XX 26-APR-2001; 2001MO-US13793.
PE
XX
XX 28-APR-2000; 2000US-0560367.
PR 14-SEP-2000; 2000US-232696P.
PR 21-DEC-2000; 2000US-257517P.
PR 03-APR-2001; 2001US-0825856.
PR 03-APR-2001; 2001US-0825876.
PR 13-APR-2001; 2001US-269020P.
XX
XX (KOSA-) KOSAN BIOSCIENCES INC.
XX
XX Arslanian RL, Ashley G, Frykman S, Julien B, Katz L, Khosla C;
PI Lau J, Licardi PJ, Regentin R, Santl D, Tang L;
XX
XX WPI; 2002-075167/10.
DR P-PSDB; ABB07167.
XX
XX Recombinant host cells useful for producing polyketides e.g. epothilone
PT or its derivatives, comprises a recombinant expression vector encoding

PT a heterologous polyketide synthase gene -
XX
XX Example 11; Page 164-165; 221pp; English.
PS
XX
XX The invention provides a recombinant host cell, of the suborder
CC Cystobacterineae comprising a recombinant expression vector encoding a
CC heterologous polyketide synthase (PKS) gene and produces a polyketide
CC synthesized by the PKS enzyme encoded on the vector. An epothilone
CC derivative of a specified formula can be produced by culturing the host
CC cell with a diketide equivalent compound of a specified formula. The host
CC cells produces polyketides at high levels and are used in the production
CC of not only epothilones, including new epothilone derivatives, but also
CC other polyketides. Methods of purifying the epothilone derivatives are
CC also useful for treating cancer, hyperproliferative diseases and
CC conditions such as psoriasis, inflammation, sarcomas, neoplasms,
CC lymphomas, multiple sclerosis, rheumatoid arthritis, atherosclerosis and
CC /or restenosis. It improves polyketide production in any organism and
CC also for production of products of recombinant PKS genes and modification
CC enzymes. The present sequence represents the nucleotide sequence of the
CC KR domain of extender module 6 of the epothilone PKS. Inactivation of
CC this domain results in a novel PKS capable of producing a 9-keto-
CC epothilone analogue.
CC XX

SQ Sequence 552 BP; 70 A; 176 C; 214 G; 92 T; 0 other;

Query Match 38.8%; Score 39.2; DB 24; Length 552;
Best Local Similarity 62.0%; Pred. No. 0.0079;
Matches 62; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY 1 CGGCGAGTGGCGCTTACAGGGCGTATCGATCGAGCGGCTCGATGCTGCTG 60
DB 233 CGTGGGGGATGCCGCTCCGCGGTGCTATCGATCGAGCGGCTGCG 292
QY 61 TTGATGACGACGACCGACCGCTTCTCGCGGTGCTG 100
DB 293 TGATGACGACGACCGCGCGGTGCTCGCGGTGCTG 332

RESULT 15
AA168064
ID AA168064 standard; DNA; 552 BP.
XX
XX AA168064;
XX
DT 13-MAR-2002 (first entry)
XX
XX 9-keto-epothilone PKS inactive KR domain nucleotide sequence.
DE
XX
XX Cystobacterinease; recombinant; polyketide synthase; PKS; polyketide;
KW epothilone; p1A gene; cytosolic; antiproliferative; antiarthritic;
KW antitartaric; antitartaric; antitartaric; neuroprotective; vasotropic; ds.
XX
XX Synthetic.
OS
FH Key Location/Qualifiers
FT CDS 1..552
FT /*tag= a
FT /note= "mutated KR domain of extender module 6"

PN WO200183800-A2.
XX
XX 08-NOV-2001.
PD
XX
XX 26-APR-2001; 2001MO-US13793.
PE
XX
XX 28-APR-2000; 2000US-0560367.
PR 14-SEP-2000; 2000US-232696P.
PR 21-DEC-2000; 2000US-257517P.
PR 03-APR-2001; 2001US-0825856.
PR 03-APR-2001; 2001US-0825876.
PR 13-APR-2001; 2001US-269020P.
XX
XX (KOSA-) KOSAN BIOSCIENCES INC.
XX

XX Arslanian RL, Ashley G, Frykman S, Julien B, Katz L, Khosia C;
PI Lau J, Licardi PJ, Regentin R, Santi D, Tang L;
XX WPI; 2002-075167/10.
DR P-PSDB; ABB07168.

XX Recombinant host cells useful for producing polyketides e.g. epothonone
PT or its derivatives, comprises a recombinant expression vector encoding
PT a heterologous polyketide synthase gene

XX Example 11; Page 165-166; 221pp; English.

XX The invention provides a recombinant host cell, of the suborder
CC Cystobacterineae comprising a recombinant expression vector encoding a
CC heterologous polyketide synthase (PKS) gene and produces a polyketide
CC synthesized by the PKS enzyme encoded on the vector. An epothonone
CC derivative of a specified formula can be produced by culturing the host
CC cell with a diketide equivalent compound of a specified formula. The host
CC cells produces polyketides at high levels and are used in the production
CC of not only epothonones, including new epothonone derivatives, but also
CC other polyketides. Methods of purifying the epothonone derivatives are
CC also useful for treating cancer, hyperproliferative diseases and
CC conditions such as psoriasis, inflammation, sarcomas, neoplasms,
CC lymphomas, multiple sclerosis, rheumatoid arthritis, atherosclerosis and
CC /or restenosis. It improves polyketide production in any organism and
CC also for production of products of recombinant PKS genes and modification
CC enzymes. The present sequence represents the nucleotide sequence of a
CC mutated and inactive KR domain of extender module 6 of the novel 9-keto-
CC epothonone PKS of the present invention.

XX Sequence 552 BP; 70 A; 177 C; 212 G; 93 T; 0 other;

Query Match 38.8%; Score 39.2; DB 24; Length 552;
Best Local Similarity 62.0%; Pred. No. 0.0079;
Matches 62; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY 1 CGGGGAGTGGCGCTTACAGGGCGGATGATCCATGACCGCGAGCGGCTGATGATGATGTC 60
DB 233 CGTGGGGATGCCCTCCGCGGCTGCTTCATGCGGCGGATTCCTGACGACGAGGCTGC 292
QY 61 TTGATGACAGACACCGACCGCTTCTCGCGGGTGCCTGGC 100
DB 293 TGATGACAGCAAAACCCCGCGGCTTCGCGGCGGATGATGCG 332

Search completed: November 6, 2002, 12:14:42
Job time : 260 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 6, 2002, 11:51:49 ; Search time 443.625 Seconds

(without alignments)
6625.829 Million cell updates/sec

Title: US-09-724-876-2_COPY_30550_30650

Perfect score: 101

Sequence: 1 cgcgcgagtcgcccgtaccag.....gtcttcgcgcggtcgtgca :01

Scoring table: IDENTITY_NNC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

```
GenBml: *
1: gb_da: *
2: gb_htg: *
3: gb_in: *
4: gb_om: *
5: gb_ov: *
6: gb_pat: *
7: gb_ph: *
8: gb_pl: *
9: gb_pr: *
10: gb_ro: *
11: gb_sts: *
12: gb_sy: *
13: gb_un: *
14: gb_vi: *
15: em_ba: *
16: em_fun: *
17: em_hum: *
18: em_in: *
19: em_mu: *
20: em_om: *
21: em_ov: *
22: em_ov: *
23: em_ph: *
24: em_pat: *
25: em_pl: *
26: em_ro: *
27: em_sts: *
28: em_un: *
29: em_vi: *
30: em_htg_hum: *
31: em_htg_inv: *
32: em_htg_other: *
33: em_htg_mus: *
34: em_htg_pln: *
35: em_htg_rtd: *
36: em_htg_mam: *
37: em_htg_vrt: *
38: em_sy: *
39: em_htgo_hum: *
40: em_htgo_mus: *
41: em_htgo_other: *
```

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------------|---------------------|
| 1 | 101 | 100.0 | 58733 | 1 AF217189 | AF217189 Sorangium |
| 2 | 101 | 100.0 | 68750 | 1 AF210843 | AF210843 Sorangium |
| 3 | 101 | 100.0 | 68750 | 6 ARI93029 | ARI93029 Sequence |
| 4 | 101 | 100.0 | 68750 | 6 ARI93029 | ARI93029 Sequence |
| 5 | 101 | 100.0 | 68750 | 6 ARI99551 | ARI99551 Sequence |
| 6 | 101 | 100.0 | 68750 | 6 ARI99559 | ARI99559 Sequence |
| 7 | 101 | 100.0 | 68750 | 6 ARI99567 | ARI99567 Sequence |
| 8 | 101 | 100.0 | 68750 | 6 AR201097 | AR201097 Sequence |
| 9 | 101 | 100.0 | 68750 | 6 AR208671 | AR208671 Sequence |
| 10 | 51.6 | 51.1 | 33529 | 6 ARI72664 | ARI72664 Sequence |
| 11 | 47.8 | 47.3 | 49736 | 1 AF319998 | AF319998 Stigmatel |
| 12 | 46.8 | 46.3 | 31422 | 6 E38021 | E38021 Avermectin |
| 13 | 46.8 | 46.3 | 64957 | 1 AB033367 | AB033367 Streptomyl |
| 14 | 44 | 43.6 | 49377 | 6 I88042 | I88042 Sequence 1 |
| 15 | 44 | 43.6 | 67523 | 1 SCU24241 | U24241 Sorangium c |
| 16 | 43.6 | 43.2 | 41097 | 1 AF016585 | AF016585 Streptomyl |
| 17 | 43.6 | 43.2 | 66808 | 1 SAU421825 | SAU421825 Stigmatel |
| 18 | 42.6 | 42.2 | 30000 | 6 AX250262 | AX250262 Sequence |
| 19 | 41.8 | 41.4 | 27522 | 1 AB070942 | AB070942 Streptomyl |
| 20 | 41.2 | 40.8 | 43280 | 1 SFU78289 | SFU78289 Streptomyl |
| 21 | 41.2 | 40.8 | 125401 | 6 AF263912 | AF263912 Streptomyl |
| 22 | 41.2 | 40.8 | 125401 | 6 AX211739 | AX211739 Sequence |
| 23 | 40.4 | 40.0 | 5088 | 6 AX153795 | AX153795 Sequence |
| 24 | 40.4 | 40.0 | 20394 | 1 SNA132222 | AX132222 Streptomyl |
| 25 | 40.4 | 40.0 | 20394 | 6 AX067996 | AX067996 Sequence |
| 26 | 40.4 | 40.0 | 24225 | 1 SC2C4 | AL512902 Streptomyl |
| 27 | 40.4 | 40.0 | 34071 | 6 AX153790 | AX153790 Sequence |
| 28 | 40.4 | 40.0 | 39314 | 1 SGR300302 | AJ300302 Streptomyl |
| 29 | 40.4 | 40.0 | 42717 | 6 AX153789 | AX153789 Sequence |
| 30 | 40.4 | 40.0 | 84985 | 1 SNA278573 | AJ278573 Streptomyl |
| 31 | 40.4 | 40.0 | 113193 | 1 AF357202 | AF357202 Streptomyl |
| 32 | 39.4 | 39.0 | 16096 | 1 AE007124 | AE007124 Mycobacte |
| 33 | 39.4 | 39.0 | 552 | 1 MTCT2461 | Z8385 Mycobacteri |
| 34 | 39.2 | 38.8 | 552 | 6 AX403009 | AX403009 Sequence |
| 35 | 39.2 | 38.8 | 552 | 6 AX403011 | AX403011 Sequence |
| 36 | 38.8 | 38.4 | 16767 | 6 AX089464 | AX089464 Sequence |
| 37 | 38.8 | 38.4 | 29736 | 6 AX089421 | AX089421 Sequence |
| 38 | 38.8 | 38.4 | 50000 | 6 AX089417 | AX089417 Sequence |
| 39 | 38.8 | 38.4 | 53789 | 1 AMM223012 | AJ223012 Amycolato |
| 40 | 38.8 | 38.4 | 53789 | 6 A69720 | A69720 Sequence 3 |
| 41 | 38.8 | 38.4 | 80161 | 1 AV007564 | AV007564 Saccharop |
| 42 | 38.8 | 38.4 | 80161 | 6 ARI65018 | ARI65018 Sequence |
| 43 | 38.8 | 38.4 | 90445 | 1 AF040570 | AF040570 Amycolato |
| 44 | 38.2 | 37.8 | 50000 | 6 AX089420 | AX089420 Sequence |
| 45 | 37.2 | 36.8 | 2700 | 6 AX211733 | AX211733 Sequence |

ALIGNMENTS

```
RESULT 1
LOCUS AF217189
DEFINITION AF217189 58733 bp DNA linear BCT 09-JUN-2000
Sorangium cellulosum putative transposase gene, partial cds;
putative transposase gene, complete cds; epothilone biosynthesis
gene cluster, complete sequence; putative membrane protein gene,
complete cds.
ACCESSION AF217189
VERSION AF217189
KEYWORDS AF217189.1 GI:7453554
SOURCE
ORGANISM Polyanium cellulosum.
Polyanimum cellulosum
Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
Myxococcales; Sorangineae; Polyaniumaceae; Polyanimum.
REFERENCE 1 (bases 1 to 58733)
```


Query Match 100.0%; Score 101; DB 6; Length 68750;
Best Local Similarity 100.0%; Pred. No. 2,4e-16;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGCGGAGTGGCCGTTACAGGCGGTGATCATGACCGGAGCGCTCATGATGCTGTC 60
|||||
Db 36162 CGCGGAGTGGCCGTTACAGGCGGTGATCATGACCGGAGCGCTCATGATGCTGTC 36221
|||||

QY 61 TTGATGAGCAGACACCGACCGCTTCTCGCGGCTGTGCA 101
|||||
Db 36222 TTGATGAGCAGACACCGACCGCTTCTCGCGGCTGTGCA 36262
|||||

RESULT 4
LOCUS AR199551 68750 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 1 from patent US 6355457.
ACCESSION AR199551
VERSION AR199551.1 GI:20249625
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and
Gorlach,J.
TITLE Genes for the biosynthesis of epoethlones
JOURNAL Patent: US 6355457-A 1 12-MAR-2002;
FEATURES
source 1..68750
location/Qualifiers
BASE COUNT 9596 a 22456 c 25539 g 11159 t
ORIGIN

Query Match 100.0%; Score 101; DB 6; Length 68750;
Best Local Similarity 100.0%; Pred. No. 2,4e-16;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGCGGAGTGGCCGTTACAGGCGGTGATCATGACCGGAGCGCTCATGATGCTGTC 60
|||||
Db 36162 CGCGGAGTGGCCGTTACAGGCGGTGATCATGACCGGAGCGCTCATGATGCTGTC 36221
|||||

QY 61 TTGATGAGCAGACACCGACCGCTTCTCGCGGCTGTGCA 101
|||||
Db 36222 TTGATGAGCAGACACCGACCGCTTCTCGCGGCTGTGCA 36262
|||||

RESULT 5
LOCUS AR199559 68750 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 1 from patent US 6355458.
ACCESSION AR199559
VERSION AR199559.1 GI:20249633
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and
Gorlach,J.
TITLE Genes for the biosynthesis of epoethlones
JOURNAL Patent: US 6355458-A 1 12-MAR-2002;
FEATURES
source 1..68750
location/Qualifiers
BASE COUNT 9596 a 22456 c 25539 g 11159 t
ORIGIN

Query Match 100.0%; Score 101; DB 6; Length 68750;
Best Local Similarity 100.0%; Pred. No. 2,4e-16;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGCGGAGTGGCCGTTACAGGCGGTGATCATGACCGGAGCGCTCATGATGCTGTC 60
|||||
Db 36162 CGCGGAGTGGCCGTTACAGGCGGTGATCATGACCGGAGCGCTCATGATGCTGTC 36221
|||||

QY 61 TTGATGAGCAGACACCGACCGCTTCTCGCGGCTGTGCA 101
|||||
Db 36222 TTGATGAGCAGACACCGACCGCTTCTCGCGGCTGTGCA 36262
|||||

|||||
Db 36162 CGCGGAGTGGCCGTTACAGGCGGTGATCATGACCGGAGCGCTCATGATGCTGTC 36221
|||||

QY 61 TTGATGAGCAGACACCGACCGCTTCTCGCGGCTGTGCA 101
|||||
Db 36222 TTGATGAGCAGACACCGACCGCTTCTCGCGGCTGTGCA 36262
|||||

RESULT 6
LOCUS AR199567 68750 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 1 from patent US 6355459.
ACCESSION AR199567
VERSION AR199567.1 GI:20249641
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and
Gorlach,J.
TITLE Genes for the biosynthesis of epoethlones
JOURNAL Patent: US 6355459-A 1 12-MAR-2002;
FEATURES
source 1..68750
location/Qualifiers
BASE COUNT 9596 a 22456 c 25539 g 11159 t
ORIGIN

Query Match 100.0%; Score 101; DB 6; Length 68750;
Best Local Similarity 100.0%; Pred. No. 2,4e-16;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGCGGAGTGGCCGTTACAGGCGGTGATCATGACCGGAGCGCTCATGATGCTGTC 60
|||||
Db 36162 CGCGGAGTGGCCGTTACAGGCGGTGATCATGACCGGAGCGCTCATGATGCTGTC 36221
|||||

QY 61 TTGATGAGCAGACACCGACCGCTTCTCGCGGCTGTGCA 101
|||||
Db 36222 TTGATGAGCAGACACCGACCGCTTCTCGCGGCTGTGCA 36262
|||||

RESULT 7
LOCUS AR201097 68750 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 1 from patent US 6358719.
ACCESSION AR201097
VERSION AR201097.1 GI:20251985
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and
Gorlach,J.
TITLE Genes for the biosynthesis of epoethlones
JOURNAL Patent: US 6358719-A 1 19-MAR-2002;
FEATURES
source 1..68750
location/Qualifiers
BASE COUNT 9596 a 22456 c 25539 g 11159 t
ORIGIN

Query Match 100.0%; Score 101; DB 6; Length 68750;
Best Local Similarity 100.0%; Pred. No. 2,4e-16;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGCGGAGTGGCCGTTACAGGCGGTGATCATGACCGGAGCGCTCATGATGCTGTC 60
|||||
Db 36162 CGCGGAGTGGCCGTTACAGGCGGTGATCATGACCGGAGCGCTCATGATGCTGTC 36221
|||||

QY 61 TTGATGAGCAGACACCGACCGCTTCTCGCGGCTGTGCA 101
|||||
Db 36222 TTGATGAGCAGACACCGACCGCTTCTCGCGGCTGTGCA 36262
|||||


```
YGPFRVAVRELYRANGALRGVQLGELTAEERYSLHALLDGCFOVLAALLEGIDS
EGGVVPAVLSRLRVLYRPEALMWSHVRSAGAGSNTIECDVTLLNDGGRVAVER
RLVCKKIAAASAPADEVDWMYISAQEIPQOARAEPTGSSRGLVLPADARGIOS
LASGLAERAGGCVVYVAGPSYRELRADRIKSPISDQALKEFSDOSPCKGMVH
LYGLDVASVAVRESEPERALSTSTSVYRLLOAMORGRDPRMLVYTCVOVIASG
EPPGSPHAPPLVGLAAVITHEPRLCAHVDLSAOPSPEELSLTEELGDGAEDRIA
LNGAVRHVARLRGISGRKAAEPRETTAGVLTITGGLGLLEVAKWTIOGARMLA
LVGRAPSEAEKAVIGLIRANGAEVAVRADVTSADVRLGLSIDSELPPLRGVVA
AGVIAVDVLTQDEPRIRAVLAPKVOGANNLVTQRELFEMVSSAAMSGSPGO
GNYAANAFMDVLAHHRRAAGLRGSLVNGMPADVGLAVATNRGRLRGVGNMP
AOALEAGRLVQPAOVAOVVPLDILROMEFEPLSAOPELSTVLMEKEYSRGRGS
ARELLAAAPBQRKGLBECYLRLEQIGRITRMPEPSRIEPPRPGDGLCMBLTELNRK
IESGLKLTLSATLTAIVPVALTYTLTGKLEPLVEAEARPAVRPRTQSPRSVGVAP

Query Match      47.3%; Score 47.8; DB 1; Length 49736;
Best Local Similarity 70.3%; Pred. No. 0.012;
Matches 64; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

QY 11 GCCGTTACAGGCGGTATCCATGACCGGAGCGCTGCATGATGTCGTGATGACGA 70
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 15263 GCCTCTGCGCGCGGTGTCATCCCGCGGCTGTCGACGATGATGTCCTCTCGACGA 15204

QY 71 GACCACGACCGCTCTCGCGGGTCTGCGCA 101
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 15203 GGACTGGAGCGCTCTCCCGGGTGTTCGA 15173

RESULT 12
E38021 31422 bp DNA linear PAT 31-JAN-2002
LOCUS Avermectin aglycon synthase gene.
DEFINITION E38021
ACCESSION E38021.1 GI:18626910
VERSION JP 2000245457-A/2.
KEYWORDS Streptomyces avermiltillis.
SOURCE Streptomyces avermiltillis.
ORGANISM Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
REFERENCE 1 (bases 1 to 31422)
AUTHORS Omura,S. and Ikeda,H.
TITLE Avermectin aglycon synthase gene
JOURNAL Patent: JP 2000245457-A 2 12-SEP-2000;
COMMENT THE KITASATO INSTITUTE
OS Streptomyces avermiltillis
PN JP 2000245457-A/2
PD 12-SEP-2000
PE 24-FEB-1999 JP 1999046961
PR
PI SATOSHI OMURA, HARUO IKEDA
PC C12N15/00,A61K31/70,C12N1/15,C12N9/88,C12P19/62, PC
C1201/88//C07D493/22,
PC (C12N1/15,C12R1:465),(C12N9/88,C12R1:465),C12N15/00 CC
FH key location/Qualifiers
FT CDS (1)-(14643)
CDS (14624)..(31419).
FEATURES
    source 1..31422
    /organism="Streptomyces avermiltillis"
    /db_xref="taxon:33903"
BASE COUNT 4136 a 10237 c 11677 g 5372 t
ORIGIN

Query Match      46.3%; Score 46.8; DB 6; Length 31422;
Best Local Similarity 67.3%; Pred. No. 0.024;
Matches 66; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 1 CGCGGAGTGGCCGTTACAGGCGGTATCCATGACCGGAGCGCTGCATGATGTCGTC 60
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 3863 CGCGGAGTGGCCGTTACAGGCGGTATCCATGACCGGAGCGCTGCATGATGTCGTC 3922

QY 61 TTGATGACGACGACGACCGCTTCGCGGGTCTGCTG 98
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 3923 TGGAGGGGTGTTCGTCGAGCGGTTCACGGGGGTCTGCTG 3960
```

```
RESULT 13
AB032367/c
LOCUS Streptomyces avermiltillis polyketide synthase gene cluster (aveA1,
DEFINITION aveA2, aveA3, aveA4) and aveC, aveB genes, complete cds.
ACCESSION AB032367
VERSION AB032367.1 GI:5902890
KEYWORDS AveC: cytochrome P450 hydroxylase; type I polyketide synthase AVES
AVES 2; type I polyketide synthase AVES 3; type I polyketide synthase
Streptomyces avermiltillis DNA.
SOURCE Streptomyces avermiltillis
ORGANISM Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
REFERENCE 1 (bases 1 to 64957)
AUTHORS Ikeda,H., Nonomiya,T., Usami,M., Ohta,T. and Omura,S.
TITLE Organization of the biosynthetic gene cluster for the polyketide
JOURNAL antimetabolic macrolide avermectin in Streptomyces avermiltillis
Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9509-9514 (1999)
9380548
REFERENCE 2 (bases 1 to 64957)
AUTHORS Ikeda,H., Nonomiya,T., Usami,M., Ohta,T. and Omura,S.
TITLE Direct Submission
JOURNAL Submitted (13-SEP-1999) Haruo Ikeda, School of Pharmaceutical
Sciences, Kitasato University, Microbial chemistry; 5-9-1
Shirokane, Minato-ku, Tokyo 108-8641, Japan
(E-mail:ikedahmc.pharm.kitasato-u.ac.jp, Tel:+81-3-5791-6242,
Fax:+81-3-3444-6197)
FEATURES
    source 1..64957
    /organism="Streptomyces avermiltillis"
    /db_xref="taxon:33903"
    101..12019
    /gene="aveA1"
    101..12019
    /note="contains a loading module, Ave module 1 and module
    2
    multifunctional polyketide synthase"
    /codon_start=1
    /transl_table=11
    /product="type I polyketide synthase AVES 1"
    /protein_id="BAA84474.1"
    /db_xref="GI:5902891"
    /translation="MORMDGGEPRPAGGEVYGVADAEADRGDVFVEFGPGQPMQMGCR
    ELLDSDVPRRESVRACEAFAYVWMSVBOYLRDSDPARGLDRVYVVPITFAVWISL
    ALMWSQVEPCAVLGHSLGETIAAHVSGGLSLADARVITLMSQOTTLAGTGLVS
    VAATPDELLPRIAPWTEENPARLAAVANGPSTVSSAGAREVADLVADLAAOVTR
    MIPVVPAPHSPLMVAIEERVVSGLPITPRPSPRIPFHSVSGRGIRDELDAAYVRN
    MSSYRFEPAARLLLOOGKRTFVEMSPHVLVWGLOELADPDGTTGADVYIMGTLR
    RGCGTLDHFTNSLAORHGERTSATVLSARLTALSPQOOSLILDIRAHVHMLND
    DCONERTADGASAFALHGPSVMGVELRNLSKATLRIRLPYTLIPRTPPAVAAAR
    LRTALGHIDEDTAVPDPSPSGHGTAAADDTALITGMACRPGGVRGSKDWEIAAS
    GDDALGPEPTDGRWPTEQRHADPTQPTFPQGGFLHDAHFAGKFFGSIPEPRLA
    MDPOORLLLETSMWEAFERAGIDPLSVRSRSGVFAGALSFDVGPMDRASSGADAVE
    GHILGTGTVSGLRISAYSGLEGPALITVDGCSALVTYLHACLSLBSGECTTLAG
    GVSVMSTLGMTEFESRGILSVDCRCAVSAADDTGMEGGLVLRISPAVGLH
    RVLAVRGSANVQDASNGCLRPNCPABEVRITROLANAGLSVANVDVVEBGCTTTL
    GDTAQLALATYTGGRADRPMLTGLSLNSNIGHTMAAGVGGLIKMVALRGVLPRT
    LHAVDRSPQVDMVSAGAVRLTEAVPMPDADAGRLRRAGVSSFGIGTNAHVLIEEAPA
    AGGCVAGGVLEGAGLALISVAEVAAPVAPAEVSPVPVPAVEEVLPGVARSAG
    LRAQAEALROYAVRVPDVSIAVDVAGLACGRAVLEBRVAVLAADEEVLQGALAAG
    EPPDRVTGTHGAPGDRGVVVFVEFGGOMGMGVRLLASSVFAARMQACEEALAPV
    YDMASVDILRRDAGAVVERADVVPVLFVSYSIALMRSGTEPDAVLGHSGEETA
    AAHVCGALSLDAAETVALRSALAAVGRGMAVSVLPAGEVEQLIGERNAGRLLWA
    AVNGPRSTAVSGDAEAVDEVILAYCAGTCVRARRIPVDASCPHYQRLREBELLAG
    ISPOSGVPEFESTVEGTWLDPTTDLDAAVYRLHQPVPFSDAVOGLADGGRLEVEVS
    PPHPLVPAIIEDETTEDTADVTAIGSLRGGDNDTRFLALAHNTHTGTGTGPTNNHY
    THNHTNPHPHHLDLPYRPFQONHWLSSQPGAGSGGACAGSAGSAGSGRAGTAGCTA
    EYESRWDVAVARODLETVATLAVPSPSGLDLVFALSAHMHQHDARINTWTQET
    WKPLTLPTTHQPHQMTWLIAIPETQTHPHHITNLIHLHHHGIITPIPLTNHTHTNPOH
```


